THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

10/506455

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

March 18, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/361,948*
FILING DATE: *March 05, 2002*
RELATED PCT APPLICATION NUMBER: *PCT/US03/03120*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. SWAIN
Certifying Officer

PRIORITY
DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c)

		Type a plus sign (+) inside this box -->
Docket Number	P-15648	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE NAME	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Beals	John	Michael	Indianapolis, Indiana
Kuchibhotla	Uma		Indianapolis, Indiana

TITLE OF THE INVENTION (280 characters max)

HETEROLOGOUS G-CSF FUSION PROTEINS

CORRESPONDENCE ADDRESS

Eli Lilly and Company
Patent Division/D.C. 1104
Lilly Corporate Center
Indianapolis, Indiana 46285

25885
PATENT TRADEMARK OFFICE

STATE	IN	ZIP CODE	46285	COUNTRY	USA

ENCLOSED APPLICATION PARTS (check all that apply)

X	Specification	Number of pages	53	
	Small Entity Statement			
X	Drawing(s)	Number of Sheets	1	
	Other (Specify)			

METHOD OF PAYMENT (check one)

	A check or money order is enclosed to cover the Provisional filing fees	**PROVISIONAL FILING FEE AMOUNT ($)**	$160.00
X	The Assistant Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 05-0840		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[X] No.

[ ] Yes, the name of the U.S. Government agency and the Government contract number are:

Respectfully submitted,
SIGNATURE _____ Date |03/05/02|

TYPED or PRINTED NAME   MARK J. STEWART

REGISTRATION NO. (if appropriate) |43,936|

[ ] Additional inventors are being named on separately numbered sheets attached hereto

PROVISIONAL APPLICATION FOR PATENT FILING ONLY

"Express Mail" mailing label number   EL018702675US
Date of Deposit   March 5, 2002
I hereby certify that this paper or fee is being deposited with the United States Postal
Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date
indicated above and is addressed to the Assistant Commissioner for Patents, Arlington, VA,
22202.

_____   _____
Printed Name   Signature

HETEROLOGOUS G-CSF FUSION PROTEINS

FIELD OF THE INVENTION

The present invention relates to heterologous fusion

5 proteins, including analogs and derivatives thereof, fused
to proteins that have the effect of extending the *in vivo*
half-life of the proteins. These fusion proteins are
significant in human medicine, particularly in the treatment
of conditions treatable by stimulation of circulating

10 neutrophils, such as after chemotherapy regimens or in
chronic congenital neutropenia. More specifically, the
invention relates to novel heterologous fusion proteins with
granulocyte-colony stimulating factor activity.

BACKGROUND OF THE INVENTION

15 Among all blood cell lineages, the modulation of
neutrophil and platelet production has been of highest
interest to clinical oncologists and hematologists.
Myelosuppression is the single most severe complication of

20 cancer chemotherapy, and a major cause of treatment delay
during multiple-cycle or combination chemotherapy. It is
also the major dose-limiting factor for most
chemotherapeutic agents. Due to the short half-lives of
neutrophils in peripheral blood, life-threatening falls in

25

30

35

"Express Mail" mailing label number __EL018702675US__

Date of Deposit __March 5, 2002__

I hereby certify that this paper or fee is being deposited with the United
States Postal Service "Express Mail Post Office to Addressee" service under 37
C.F.R. 1.10 on the date indicated above and is addressed to the Assistant
Commissioner for Patents, Washington, D.C. 20231.

_____   _____
Printed Name   Signature

neutrophil levels are seen after a number of conventional anti-tumor chemotherapy regimens.

5   The most prominent regulator of granulopoiesis is granulocyte-colony stimulating factor (G-CSF). G-CSF induces proliferation and differentiation of hematopoietic progenitor cells resulting in increased numbers of circulating neutrophils. G-CSF also stimulates the release of mature neutrophils from bone marrow and activates their 10   functional state. [Souza L.M., et al. (1986) Science 232:61-65]. Thus, therapeutic proteins with G-CSF activity have tremendous value in situations where there are reduced circulating levels of neutrophilic granuloctyes.

         However, the usefulness of therapy using G-CSF peptides 15   has been limited by their short plasma half-life. Thus, they must be administered intravenously or subcutaneously at fairly frequent intervals (once or twice a day) in order to maintain their neutrophil stimulating properties. In addition, this short half-life limits the performance of the 20   drug to traditional drug delivery systems. It would clearly benefit the treatment of patients with abnormally low neutrophils, and reduce the discomfort and inconvenience associated with frequent injections to provide a pharmaceutical agent that could be administered less 25   frequently and optionally by alternative routes of administration. Thus, a need exists to develop agents that stimulate the production of mature neutrophils and are more optimal in their duration of effect.

         The present invention overcomes the problems associated 30   with delivering a compound that has a short plasma half-life in two respects. First, G-CSF is hyperglycosylated. The carbohydrate content of G-CSF is altered by substituting amino acids that can act as substrates for glycosylating enzymes in mammalian cells. Most significantly, the present 35   invention encompasses fusion of these hyperglycosylated G-

CSF analogs to another protein with a long circulating half-life such as the Fc portion of an immunoglobulin or albumin.

BRIEF SUMMARY OF THE INVENTION

5   Compounds of the present invention include heterologous fusion proteins comprising a hyperglycosylated G-CSF analog fused to a polypeptide selected from the group consisting of

a) human albumin;

b) human albumin analogs; and

10   c) fragments of human albumin.

Additional compounds of the present invention include a heterologous fusion protein comprising a hyperglycosylated G-CSF analog fused to a polypeptide selected from the group consisting of

15   a) the Fc portion of an immunoglobulin;

b) an analog of the Fc portion of an immunoglobulin; and

c) fragments of the Fc portion of an immunoglobulin.

The hyperglycosylated G-CSF analog may be fused to the

20   polypeptide via a peptide linker. It is preferable that the peptide linker is selected from the group consisting of:

a) a glycine rich peptide;

b) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]$_n$ where n is 1, 2, 3, 4, or 5; and

25   c) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]$_3$

One aspect of the present invention includes heterologous fusion proteins, wherein the hyperglycosylated

30   G-CSF analogs have the Formula (I) [SEQ ID NO:1]

```
·1                5                    10                    15
Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
              20                   25                   30
35   Xaa Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
              35                   40                   45
     Glu Lys Leu Cys Xaa Xaa Xaa Lys Leu Cys His Pro Glu Glu Leu Val
              50                   55                   60
     Leu Leu Gly His Ser Leu Gly Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
```

```
         65                      70                      75                      80
         Xaa Xaa Xaa Xaa Xaa Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
                     85                      90              '95
         Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Xaa Xaa Xaa Ser
     5          100                     105             110
         Xaa Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
             115                     120             125
         Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
             130                     135             140
    10   Ala Leu Gln Pro Xaa Xaa Xaa Ala Met Pro Ala Phe Xaa Xaa Xaa Phe
         145                 150             155             160
         Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
                     165             170
         Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro      (I)
    15
```

wherein:
Xaa at position 17 is Cys, Ala, Leu, Ser, or Glu;
Xaa at position 37 is Ala or Asn;
Xaa at position 38 is Thr, or any other amino acid except
20   Pro;
Xaa at position 39 is Tyr, Thr, or Ser;
Xaa at position 57 is Pro or Val;
Xaa at position 58 is Trp or Asn;
Xaa at position 59 is Ala or any other amino acid except
25   Pro;
Xaa at position 60 is Pro, Thr, Asn, or Ser,
Xaa at position 61 is Leu, or any other amino acid except
     Pro;
Xaa at position 62 is Ser or Thr;
30   Xaa at position 63 is Ser or Asn;
Xaa at position 64 is Cys or any other amino acid except
     Pro;
Xaa at position 65 is Pro, Ser, or Thr;
Xaa at position 66 is Ser or Thr;
35   Xaa at position 67 is Gln or Asn;
Xaa at position 68 is Ala or any other amino acid except
     Pro;
Xaa at position 69 is Leu, Thr, or Ser
Xaa at position 93 is Glu or Asn
40   Xaa at position 94 is Gly or any other amino acid except
     Pro;
Xaa at position 95 is Ile, Asn, Ser, or Thr;
Xaa at position 97 is Pro, Ser, Thr, or Asn;
Xaa at position 133 is Thr or Asn;
45   Xaa at position 134 is Gln or any other amino acid except
     Pro;
Xaa at position 135 is Gly, Ser, or Thr
Xaa at position 141 is Ala or Asn;
Xaa at position 142 is Ser or any other amino acid except
50   Pro; and

Xaa at position 143 is Ala, Ser, or Thr;

and wherein:
Xaa at positions 37, 38, and 39 constitute region 1;
5   Xaa at positions 58, 59, and 60 constitute region 2;
Xaa at positions 59, 60, and 61 constitute region 3;
Xaa at positions 60, 61, and 62 constitute region 4;
Xaa at positions 61, 62, and 63 constitute region 5;
Xaa at positions 62, 63, and 64 constitute region 6;
10   Xaa at positions 63, 64, and 65 constitute region 7;
Xaa at positions 64, 65, and 66 constitute region 8;
Xaa at positions 67, 68, and 69 constitute region 9;
Xaa at positions 93, 94, and 95 constitute region 10;
Xaa at positions 94, 95, and Ser at position 96
15   constitute region 11;
Xaa at positions 95, and 97, and Ser at position 96
        constitute region 12;
Xaa at positions 133, 134, and 135 constitute
        region 13;
20   Xaa at positions 141, 142, and 143 constitute
        region 14;

and provided that at least one of regions 1 through 14 comprises the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any
25   amino acid except Pro and Xaa2 is Ser or Thr.

Thus, the heterologous fusion proteins of the present invention include analogs wherein one or any combination of two or more regions comprise the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and Xaa2 is Ser or
30   Thr.

Preferred hyperglycosylated G-CSF analogs that make up part of the heterologous fusion proteins of the present invention, include the following:
35   a) G-CSF[A37N,Y39T]

b) G-CSF[P57V,W58N,P60T]

c) G-CSF[P60N,S62T]

d) G-CSF[S63N,P65T]

e) G-CSF[Q67N,L69T]

5   f) G-CSF[E93N,I95T]

g) G-CSF[T133N,G135T]

h) G-CSF[A141N,A143T]

i) G-CSF[A37N,Y39T,P57V,W58N,P60T]

j) G-CSF[A37N,Y39T,P60N,S62T]

10   k) G-CSF[A37N,Y39T,S63N,P65T]

l) G-CSF[A37N,Y39T,Q67N,L69T]

m) G-CSF[A37N,Y39T,E93N,I95T]

n) G-CSF[A37N,Y39T,T133N,G135T]

o) G-CSF[A37N,Y39T,A141N,A143T]

15   p) G-CSF[A37N,Y39T,P57V,W58N,P60T,S63N,P65T]

q) G-CSF[A37N,Y39T,P57V,W58N,P60T,Q67N,L69T]

r) G-CSF[A37N,Y39T,S63N,P65T,E93N,I95T]

The present invention also includes heterologous fusion
proteins, which are the product of the expression in a host
20   cell of an exogenous DNA sequence, which comprises a DNA
sequence encoding a heterologous fusion protein of Formula I
(described above) fused to a DNA sequence encoding human
albumin or the Fc portion of an immunolglobulin.

A hyperglcosylated heterologous fusion protein of the
25   present invention also includes polynucleotides encoding the
heterologous fusion protein described herein, vectors
comprising these polynucleotides and host cells transfected
or transformed with the vectors described herein. Also
included is a process for producing a heterologous fusion
30   protein comprising the steps of transcribing and translating
a polynucleotide described herein under conditions wherein
the heterologous fusion protein is expressed in detectable
amounts.

The present invention encompasses a method for
increasing neutrophil levels in a mammal comprising the
administration of a therapeutically effective amount of a
heterologous fusion protein described above. The present
5   invention also includes the use of the heterologous fusion
proteins described above for the manufacture of a medicament
for the treatment of patients with insufficient circulating
neutrophil levels.

The present invention also encompasses a pharmaceutical
10   formulation adapted for the treatment of patients with
insufficient neutrophil levels comprising a glycosylated
protein as described above.

BRIEF DESCRIPTION OF THE DRAWINGS

15   The invention is further illustrated with reference to
the following drawings:

Figure 1: Schematic illustrating fourteen regions in
human G-CSF wherein the amino acid sequence can be mutated
to create functional glycosylation sites.

20

DETAILED DESCRIPTION OF THE INVENTION

The present invention comprises heterologous fusion
proteins. As used herein, the term heterologous fusion
protein means a hyperglycosylated G-CSF analog fused to
25   human albumin, a human albumin analog, a human albumin
fragment, the Fc portion of an immunoglobulin, an analog of
the Fc portion of an immunoglobulin, or a fragment of the Fc
portion of an immunoglobulin. The G-CSF analog may be fused
directly, or fused via a peptide linker, to an albumin or Fc
30   protein. The albumin and Fc portion may be fused to the G-
CSF analogs at either terminus or at both termini. These
heterologous fusion proteins are biologically active and
have an increased half-life compared to native G-CSF.

Hyperglycosylated G-CSF Analogs

Hyperglycosylated analogs of G-CSF as used herein refers to analogs of G-CSF with one or more changes in the amino acid sequence which result in an increase in the

5 number of sites for carbohydrate attachment compared with native human G-CSF expressed in animal cells *in vivo*. In addition, hyperglycosylated G-CSF analogs include human G-CSF wherein the O-linked glycosylation site at position 133 is replaced with an N-linked glycosylation site. Analogs

10 are generated by site directed mutagenesis having substitution of amino acid residues creating new sites that are available for glycosylation. G-CSF analogs having a greater carbohydrate content than that found in native human G-CSF are generated by adding glycosylation sites that do

15 not perturb the secondary, tertiary, and quaternary structure required for activity. The heterologous fusion proteins of the present invention thus have a larger mass and an increased negative charge compared to native G-CSF, they will not be as rapidly cleared from the circulation.

20 It is preferred that the hyperglycosylated G-CSF analog portion of the heterologous fusion protein have 1, 2, 3, or 4 additional sites for N-glycosylation. Figure 1 illustrates fourteen different regions that can be glycosylated with very little effect on *in vitro* activity.

25 Each region may be mutated to the consensus site for N-glycosylation addition which is Asn X1 X2 wherein X1 is any amino acid except Pro and X2 is Ser or Thr. It is preferred that the X1 amino acid be any other amino acid except Trp, Asp, Glu, or Leu and it is most preferred that the X1 amino

30 acid be the naturally occurring amino acid. The scope of the present invention includes analogs wherein a single region (1 through 14) is mutated or wherein a region is mutated in combination with one or more other regions.

A representative number of fusion proteins having

35 single of multiple additional glycosylation sites have been expressed, purified, and characterized. The mutated amino

acids in the hyperglycosylated G-CSF portion of the
heterologous fusion protein are identified in parentheses.
For example, G-CSF[A37N,Y39T] is G-CSF wherein the amino
acids at positions 37 and 39 have been substituted to create

5   a glycosylation site. This site of carbohydrate attachment
is illustrated as region 1 in Figure 1. G-
CSF[A37N,Y39T,P57V,W58N,P60T] is an example of a G-CSF
analog wherein amino acids in region 1 and region 2 are
mutated to provide two functional glycosylation sites on a

10   single molecule (Figure 1).

G-CSF[A37N,Y39T,P57V,W58N,P60T,Q67N,L69T] is an example
of a G-CSF analog wherein the amino acids in region 1,
region 2, and region 9 are mutated to provide three
functional glycosylation sites on a single molecule (Figure

15   1).

Native G-CSF can be used as the backbone to create the
hyperglycosylated G-CSF analog portion of a particular
heterologous fusion protein. In addition, the native G-CSF
backbone can be modified such that substitutions in the

20   regions defined in Figure 1 are made in the context of a
different or improved heterologous fusion protein. For
example, native G-CSF with a Cysteine to Alanine
substitution at position 17 may reduce aggregation and
enhance stability and thus, can be used as the backbone used

25   to create the hyperglycosylated G-CSF analog portion of the
heterologous fusion proteins of the present invention.

In addition, Reidhaar-Olson, *et al.*, through alanine
scanning mutagenesis, describe residues critical to the
activity of human G-CSF. [Reidhaar-Olson, *et al.* (1996)

30   *Biochemistry* 35:9034-9041; See also Young, *et al.* (1997)
Protein Science 6:1228-1236]. Thus, the hyperglycosylated
G-CSF analogs can be modified by substituting amino acids
outside the glycosylated regions described in Figure 1.

The heterologous fusion proteins of the present invention also encompass G-CSF analogs wherein the O-linked glycosylation site at position 133 is mutated to serve as an N-linked glycosylation site. The N-linked carbohydrate will

5 generally have a higher sialic acid content which will protect it from the rapid clearance mechanisms associated with native G-CSF.

The functions of a carbohydrate chain greatly depend on the structure of the attached carbohydrate moiety.

10 Typically compounds with a higher sialic acid content will have better stability and longer half-lives *in vivo*. The N-linked oligosaccharides contain sialic acid in both an $\alpha2,3$ and an $\alpha2,6$ linkage to galactose. [Takeuchi, *et al.* (1988) *J. Biol. Chem.* 263:3657]. Typically the sialic acid in the

15 $\alpha2,3$ linkage is added to galactose on the mannose $\alpha1,6$ branch and the sialic acid in the $\alpha2,6$ linkage is added to the galactose on the mannose $\alpha1,3$ branch. The enzymes that add these sialic acids (β-galactoside $\alpha2,3$ sialyltransferase and β-galactoside $\alpha2,6$ sialyltransferase) are most efficeint

20 at adding sialic acid to the mannose $\alpha1,6$ and mannose $\alpha1,3$ branches respectively.

Tetra-antennary N-linked oligosachharides most commonly provide four possible sites for sialic acid attachment while bi- and tri-antennary oligosaccharide chains, which can

25 substitute for the tetra-antennary form at Asn-liked sites, commonly have at most only two or three sialic acids attached. O-linked oligosaccharides commonly provide only two sites for sialic acid attachement. Mammalian cell cultures can be screened for those cells that preferentially

30 add teta-antennary chains to the G-CSF analogs of the present invention, thereby maximizing the number of sites for sialic acid attachment. Different types of mammalian cells also differ with respect to the transferase enzymes present and consequently the sialic acid content and type of

35 oligosachharide attached at each site. One way to optimize the carbohydrate content for a given G-CSF analog is to

express the analog in a cell line wherein an expression plasmid containing DNA encoding a specific sialyl transferase (e.g., α2,6 sialyltrasnferase) is co-transfected with the G-CSF analog expression plasmid.

5 Alternatively a host cell line may be stably transfected with a sialyl transferase cDNA and that host cell used to express the G-CSF analog of interest. Thus, it is preferable if the oligosaccharide structure and sialic acid content are optimized for each analog encompassed by the

10 present invention.

Heterologous Fc fusion proteins:

The hyperglycosylated G-CSF analogs described above can be fused directly or via a peptide linker to the Fc portion of an immunoglobulin. For example SEQ ID NO:23 represents

15 the sequence of a G-CSF-IgG1 Fc fusion protein, SEQ ID NO:24 represents the sequence of a G-CSF-IgG4 Fc fusion protein, and SEQ ID NO:25 represents the sequence of a G-CSF-human albumin fusion protein.

20 Immunoglobulins are molecules containing polypeptide chains held together by disulfide bonds typically having two light chains and two heavy chains. In each chain, one domain (V) has a variable amino acid sequence depending on the antibody specificity of the molecule. The other domains

25 (C) have a rather constant sequence common to molecules of the same class.

As used herein, the Fc portion of an immunoglobulin has the meaning commonly given to the term in the field of immunology. Specifically, this term refers to an antibody

30 fragment which is obtained by removing the two antigen binding regions (the Fab fragments) from the antibody. Thus, the Fc portion is formed from approximately equal sized fragments of the constant region from both heavy chains, which associate through non-covalent interactions

35 and disulfide bonds. The Fc portion can include the hinge regions and extend through the CH2 and CH3 domains to the C-

terminus of the antibody. For example, SEQ ID NOs: 18 and 19
represent the sequence of a Fc region encompassing the hinge
region as well as the CH2 and CH3 domains. Other
representative hinge regions for human and mouse

5   immunoglobulins can be found in Antibody Engineering, A
Practical Guide, Borrebaeck, C.A.K., ed., W.H. Freeman and
Co., 1992, the teachings of which are herein incorporated by
reference.

There are five types of human immunoglobulin Fc regions

10   with different effector and pharmacokinetic properties: IgG,
IgA, IgM, IgD, and IgE. IgG is the most abundant
immunoglobulin in serum. IgG also has the longest half-life
in serum of any immunoglobulin (23 days). Unlike other
immunoglobulins, IgG is efficiently recirculated following

15   binding to an Fc receptor. There are four IgG subclasses
G1, G2, G3, and G4, each of which have different effector
functions. G1, G2, and G3 can bind C1q and fix complement
while G4 cannot. Even though G3 is able to bind C1q more
efficiently than G1, G1 is more effective at mediating

20   complement-directed cell lysis. G2 fixes complement very
inefficiently. The C1q binding site in IgG is located at
the carboxy terminal region of the CH2 domain.

All IgG subclasses are capable of binding to Fc
receptors (CD16, CD32, CD64) with G1 and G3 being more

25   effective than G2 and G4. The Fc receptor-binding region of
IgG is formed by residues located in both the hinge and the
carboxy terminal regions of the CH2 domain.

IgA can exist both in a monomeric and dimeric form held
together by a J-chain. IgA is the second most abundant Ig

30   in serum, but it has a half-life of only 6 days. IgA has
three effector functions. It binds to an IgA specific
receptor on macrophages and eosinophils, which drives
phagocytosis and degranulation, respectively. It can also
fix complement via an unknown alternative pathway.

35   IgM is expressed as either a pentamer or a hexamer,
both of which are held together by a J-chain. IgM has a

serum half-life of 5 days. It binds weakly to Clq via a binding site located in its CH3 domain. IgD has a half-life of 3 days in serum. It is unclear what effector functions are attributable to this Ig. IgE is a monomeric Ig and has

5   a serum half-life of 2.5 days. IgE binds to two Fc receptors which drives degranulation and results in the release of pro-inflammatory agents.

Depending on the desired *in vivo* effect, the heterologous fusion proteins of the present invention may

10   contain any of the isotypes described above or may contain mutated Fc regions wherein the complement and/or Fc receptor binding functions have been altered. For example, one embodiment of the present invention is a heterologous fusion protein wherein the Fc portion comprises a human IgG4

15   sequence wherein serine at position 229 is changed to proline which reduces the effector function of the protein. (See SEQ ID NO: 22).

The heterologous fusion proteins of the present invention may contain the entire Fc portion of an

20   immunoglobulin, fragments of the Fc portion of an immunoglobulin, or analogs thereof fused to a G-CSF analog. Furthermore, the Fc portion may be fused at either terminus or at both termini.

The heterologous fusion proteins of the present

25   invention can consist of single chain proteins or as multi-chain polypeptides. Two or more Fc fusion proteins can be produced such that they interact through disulfide bonds that naturally form between Fc regions. These multimers can be homogeneous with respect to the G-CSF analog or they may

30   contain different G-CSF analogs fused at the N-terminus of the Fc portion of the fusion protein.

Regardless of the final structure of the fusion protein, the Fc or Fc-like region serves to prolong the in vivo plasma half-life of the G-CSF analog compared to native

35   G-CSF. Furthermore, the fused G-CSF analog must retain some

biological activity. Biological activity can be determined by in vitro and in vivo methods known in the art.

Since the Fc region of IgG produced by proteolysis has the same in vivo half-life as the intact IgG molecule and
5   Fab fragments are rapidly degraded, it is believed that the relevant sequence for prolonging half-life resides in the CH2 and/or CH3 domains. Further, it has been shown in the literature that the catabolic rates of IgG variants that do not bind the high-affinity Fc receptor or C1q are
10 indistinguishable from the rate of clearance of the parent wild-type antibody, indicating that the catabolic site is distinct from the sites involved in Fc receptor or C1q binding. [Wawrzynczak et al., (1992) *Molecular Immunology* 29:221]. Site-directed mutagenesis studies using a murine
15 IgG1 Fc region suggested that the site of the IgG1 Fc region that controls the catabolic rate is located at the CH2-CH3 domain interface.

Based on these studies, Fc regions can be modified at the catabolic site to optimize the half-life of the fusion
20 proteins. It is preferable that the Fc region used for the heterologous fusion proteins of the present invention be derived from an IgG1 such as that represented by SEQ ID NO: 21 or an IgG4 Fc region such as that represented by SEQ ID NO: 22. It is even more preferable that the Fc region be
25 IgG4 or derived from IgG4. Preferably the IgG Fc region contains both the CH2 and CH3 regions including the hinge region.

Heterologous albumin fusion proteins:
30 The G-CSF analogs described above can be fused directly or via a peptide linker to albumin or an analog, fragment, or derivative thereof. SEQ ID NO:25 provides a representative G-CSF-human albumin fusion protein.

Generally the albumin proteins making up part of the
35 heterologous fusion proteins of the present invention can be derived from albumin cloned from any species. However,

human albumin and fragments and analogs thereof are preferred to reduce the risk of the fusion protein being immunogenic in humans. Human serum albumin (HA) consists of a single non-glycosylated polypeptide chain of 585 amino

5　acids with a formula molecular weight of 66,500. The amino acid sequence of human albumin is represented as SEQ ID NO: 20. [See Meloun, et al. (1975) FEBS Letters 58:136; Behrens, et al. (1975) Fed. Proc. 34:591; Lawn, et al. (1981) Nucleic Acids Research 9:6102-6114; Minghetti, et al.

10　(1986) J. Biol. Chem. 261:6747]. A variety of polymorphic variants as well as analogs and fragments of albumin have been described. [See Weitkamp, et al., (1973) Ann. Hum. Genet. 37:219]. For example, in EP 322,094, the inventors disclose various shorter forms of HA. Some of these

15　fragments include HA(1-373), HA(1-388), HA(1-389), HA(1-369), and HA(1-419) and fragments between 1-369 and 1-419. EP 399,666 discloses albumin fragments that include HA(1-177) and HA(1-200) and fragments between HA(1-177) and HA(1-200).

20　　　It is understood that the heterologous fusion proteins of the present invention include hyperglycosylated G-CSF analogs that are coupled to any albumin protein including fragments, analogs, and derivatives wherein such fusion protein is biologically active and has a longer plasma half-

25　life than native G-CSF. Thus, the albumin portion of the fusion protein need not necessarily have a plasma half-life equal to that of native human albumin. In addition, the albumin may be fused to either terminus or both termini of the hyperglycosylated G-CSF analog.

30　　　The heterologous fusion proteins of the present invention encompass proteins having conservative amino acid substitutions in the hyperglycosylated G-CSF analog and/or the Fc or albumin portion of the fusion protein.

　　　The term "amino acid" is used herein in its broadest

35　sense, and includes naturally occurring amino acids as well as non-naturally occurring amino acids, including amino acid

analogs and derivatives. The latter includes molecules containing an amino acid moiety. One skilled in the art will recognize, in view of this broad definition, that reference herein to an amino acid includes, for example,

5 naturally occurring proteogenic L-amino acids; D-amino acids; chemically modified amino acids such as amino acid analogs and derivatives; naturally occurring non-proteogenic amino acids such as norleucine, β-alanine, ornithine, GABA, etc.; and chemically synthesized compounds having properties

10 known in the art to be characteristic of amino acids. As used herein, the term "proteogenic" indicates that the amino acid can be incorporated into a peptide, polypeptide, or protein in a cell through a metabolic pathway.

The incorporation of non-natural amino acids, including

15 synthetic non-native amino acids, substituted amino acids, or one or more D-amino acids into the heterologous fusion proteins of the present invention can be advantageous in a number of different ways. D-amino acid-containing peptides, etc., exhibit increased stability *in vitro* or *in vivo*

20 compared to L-amino acid-containing counterparts. Thus, the construction of peptides, etc., incorporating D-amino acids can be particularly useful when greater intracellular stability is desired or required. More specifically, D-peptides, etc., are resistant to endogenous peptidases and

25 proteases, thereby providing improved bioavailability of the molecule, and prolonged lifetimes *in vivo* when such properties are desirable. Additionally, D-peptides, etc., cannot be processed efficiently for major histocompatibility complex class II-restricted presentation to T helper cells,

30 and are therefore less likely to induce humoral immune responses in the whole organism.

Although the heterologous fusion proteins of the present invention can be made by a variety of different methods, recombinant methods are preferred.

35

Construction of DNA encoding the heterologous fusion
proteins of the present invention:

Wild type albumin and immunoglobulin proteins can be
obtained from a variety of sources. For example, these
5   proteins can be obtained from a cDNA library prepared from
tissue or cells which express the mRNA of interest at a
detectable level. Libraries can be screened with probes
designed using the published DNA or protein sequence for the
particular protein of interest.

10   Screening a cDNA or genomic library with the selected
probe may be conducted using standard procedures, such as
described in Sambrook et al., *Molecular Cloning: A*
Laboratory Manual, Cold Spring Harbor Laboratory Press, NY
(1989). An alternative means to isolate a gene encoding an
15   albumin or immunoglobulin protein is to use PCR methodology
[Sambrook et al., supra; Dieffenbach et al., *PCR Primer: A*
Laboratory Manual, Cold Spring Harbor Laboratory Press, NY
(1995)]. PCR primers can be designed based on published
sequences.

20   Generally the full-length wild-type sequences cloned
from a particular species can serve as a template to create
analogs, fragments, and derivatives that retain the ability
to confer a longer plasma half-life on the G-CSF analog that
is part of the fusion protein. It is preferred that the Fc
25   and albumin portions of the heterologous fusion proteins of
the present invention be derived from the native human
sequence in order to reduce the risk of potential
immunogenicity of the fusion protein in humans.

In particular, it is preferred that the immunoglobulin
30   portion of a fusion protein encompassed by the present
invention contain only an Fc fragment of the immunoglobulin.
Depending on whether particular effector functions are
desired and the structural characteristics of the fusion
protein, an Fc fragment may contain the hinge region along
35   with the CH2 and CH3 domains or some other combination
thereof. These Fc fragments can be generated using PCR

techniques with primers designed to hybridize to sequences corresponding to the desired ends of the fragment. Similarly, if fragments of albumin are desired, PCR primers can be designed which are complementary to internal albumin

5   sequences. PCR primers can also be designed to create restriction enzyme sites to facilitate cloning into expression vectors.

DNA encoding human G-CSF can be obtained from a cDNA library prepared from tissue or cells which express G-CSF

10   mRNA at a detectable level such as monocytes, macrophages, vascular endothelial cells, fibroblasts, and some human malignant and leukemic myeloblastic cells. Libraries can be screened with probes designed using the published DNA sequence for human G-CSF. [Souza L., *et al.* (1986) *Science*

15   232:61-65]. Screening a cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, NY (1989). An alternative means to isolate the gene encoding

20   human G-CSF is to use PCR methodology [Sambrook, *et al.,* supra; Dieffenbach, *et al., PCR Primer: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, NY (1995)].

The glycosylated G-CSF analogs of the present invention can be constructed by a variety of mutagenesis

25 · techniques well known in the art. Specifically, a representative number of glycosylated G-CSF analogs were constructed using mutagenic PCR from a cloned wild-type human G-CSF DNA template (Example 1).

The glycosylated G-CSF analogs of the present invention

30   may be produced by other methods including recombinant DNA technology or well known chemical procedures, such as solution or solid-phase peptide synthesis, or semi-synthesis in solution beginning with protein fragments coupled through conventional solution methods.

35   Recombinant DNA methods are preferred for producing the glycosylated G-CSF analogs of the present invention. Host

cells are transfected or transformed with expression or cloning vectors described herein for glycosylated G-CSF analog production and cultured in conventional nutrient media modified as appropriate for inducing promoters,

5 selecting transformants, or amplifying the genes encoding the desired sequences (Example 2). The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation

Physical stability is an essential feature for

10 therapeutic formulations. The physical stability of the heterologous fusion proteins of the present invention depends on their conformational stability, the number of charged residues (pI of the protein), the ionic strength and pH of the formulation, and the protein concentration, among

15 other possible factors. As discussed previously, the G-CSF analog portion of the heterologous fusion proteins can be successfully glycosylated and expressed such that the three dimensional structure is maintained. Because these heterologous fusion proteins are able to fold properly in a

20 hyperglycosylated state, they will have improved conformational and physical stability relative to wild-type G-CSF.

While wild-type G-CSF produced in mammalian cells and bacterial cells has similar activity *in vivo*, the mammalian

25 cell-produced protein has increased conformational and physical stability due to the presence of a single O-linked sugar moiety present at position 133. Thus, the G-CSF analog portion of the heterologus fusion proteins, which has an increased glycosylation content compared to wild-type G-

30 CSF produced in mammalian or bacterial cells, will confer increased stability on the heterologous fusion protein. Furthermore, it is likely that glycosylation may inhibit inter-domain interactions and consequently enhance stability by preventing inter-domain disulfide shuffling.

35 The gene encoding a heterologous fusion protein can be constructed by ligating DNA encoding a G-CSF analog in-frame

to DNA encoding an albumin or Fc protein. The gene encoding the G-CSF analog and the gene encoding the albumin or Fc protein can also be joined in-frame via DNA encoding a linker peptide.

5   The *in vivo* function and stability of the heterologous fusion proteins of the present invention can be optimized by adding small peptide linkers to prevent potentially unwanted domain interactions. Although these linkers can potentially be any length and consist of any combination of amino acids, 10 it is preferred that the length be no longer than necessary to prevent unwanted domain interactions and/or optimize biological activity and/or stability. Generally, the linkers should not contain amino acids with extremely bulky side chains or amino acids likely to introduce significant 15 secondary structure. It is preferred that the linker be serine-glycine rich and be less than 30 amino acids in length. It is more preferred that the linker be no more than 20 amino acids in length. It is even more preferred that the linker be no more than 15 amino acids in length. A 20 preferred linker contains repeats of the sequence Gly-Gly-Gly-Gly-Ser. It is preferred that there be between 2 and 6 repeats of this sequence. It is even more preferred that there be between 3 and 4 repeats of this sequence.

To construct the heterolgous G-CSF fusion proteins, the 25 DNA encoding wild-type G-CSF, albumin, and Fc polypeptides and fragments thereof can be mutated either before ligation or in the context of a cDNA encoding an entire fusion protein. A variety of mutagenesis techniques are well known in the art. For example, a mutagenic PCR method utilizes 30 strand overlap extension to create specific base mutations for the purposes of changing a specific amino acid sequence in the corresponding protein. This PCR mutagenesis requires the use of four primers, two in the forward orientation (primers A and C) and two in the reverse orientation 35 (primers B and D). A mutated gene is amplified from the wild-type template in two different stages. The first

reaction amplifies the gene in halves by performing an A to
B reaction and a separate C to D reaction wherein the B and
C primers target the area of the gene to be mutated. When
aligning these primers with the target area, they contain
5   mismatches for the bases that are targeted to be changed.
Once the A to B and C to D reactions are complete, the
reaction products are isolated and mixed for use as the
template for the A to D reaction. This reaction then yields
the full, mutated product.
10   Once a gene encoding an entire fusion protein is
produced it can be cloned into an appropriate expression
vector. Specific strategies that can be employed to make
the G-CSF fusion proteins of the present invention are
described in example 1.
15

General methods to recombinantly express the heterologous
fusion proteins of the present invention:
        Host cells are transfected or transformed with expression
or cloning vectors described herein for heterologous fusion
20   protein production and cultured in conventional nutrient media
modified as appropriate for inducing promoters, selecting
transformants, or amplifying the genes encoding the desired
sequences. The culture conditions, such as media,
temperature, pH and the like, can be selected by the skilled
25   artisan without undue experimentation. In general,
principles, protocols, and practical techniques for maximizing
the productivity of cell cultures can be found in *Mammalian*
Cell Biotechnology: A Practical Approach, M. Butler, ed. (IRL
Press, 1991) and Sambrook, *et al.*, supra. Methods of
30   transfection are known to the ordinarily skilled artisan, for
example, $CaPO_4$ and electroporation. General aspects of
mammalian cell host system transformations have been described
in U.S. Patent No. 4,399,216. Transformations into yeast are
typically carried out according to the method of van Solingen,
35   *et al.*, *J Bact.* 130(2): 946-7 (1977) and Hsiao, *et al.*, *Proc.*
Natl. Acad. Sci. USA 76(8): 3829-33 (1979). Suitable host

cells for the expression of the fusion proteins of the present invention are derived from multicellular organisms.

The fusion proteins of the present invention may be recombinantly produced directly, or as a protein having a
5 signal sequence or other additional sequences which create a specific cleavage site at the N-terminus of the mature fusion protein. In general, the signal sequence may be a component of the vector, or it may be a part of the fusion protein-encoding DNA that is inserted into the vector. The signal
10 sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including
15 Saccharomyces and Kluyveromyces cc-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the C. albicans glucoamylase leader (EP 362,179), or the signal described in WO 90/13646. In mammalian cell expression, mammalian signal sequences may be used to direct
20 secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species as well as viral secretory leaders.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells
25 competent to take up the fusion protein-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described [Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77(7): 4216-20
30 (1980)]. A suitable selection gene for use in yeast is the trp1 gene present in the yeast plasmid Yrp7 [Stinchcomb, *et al., Nature* 282(5734): 39-43 (1979); Kingsman, *et al., Gene* 7(2): 141-52 (1979); Tschumper, *et al., Gene* 10(2): 157-66 (1980)]. The trp1 gene provides a selection marker for a
35 mutant strain of yeast lacking the ability to grow in

tryptophan, for example, ATCC No. 44076 or PEPC1 [Jones, *Genetics* 85: 23-33 (1977)].

   Various forms of a fusion protein may be recovered from culture medium or from host cell lysates. If membrane-bound,
5   it can be released from the membrane using a suitable detergent solution (e.g., Triton-X 100) or by enzymatic cleavage. Cells employed in expression of a fusion protein can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or
10   cell lysing agents.

Purification of the heterologous fusion proteins of the present invention:

   Once the heterologous fusion proteins of the present
15   invention are expressed in the appropriate host cell, the analogs can be isolated and purified. The following procedures are exemplary of suitable purification procedures:

   Various methods of protein purification may be employed and such methods are known in the art and described, for
20   example, in Deutscher, *Methods in Enzymology* 182: 83-9 (1990) and Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag, NY (1982). The purification step(s) selected will depend on the nature of the production process used and the particular fusion protein produced. For example, fusion
25   proteins comprising an Fc fragment can be effectively purified using a Protein A or Protein G affinity matix. Low or high pH buffers can be used to elute the fusion protein from the affinity matrix. Mild elution conditions will aid in preventing irreversible denaturation of the fusion protein.
30   Imidazole-containing buffers can also be used. Example 3 describes some successful purification protocols for the fusion proteins of the present invention.

Characterization of the heterologous fusion proteins of the
35   *present invention:*

Numerous methods exist to characterize the fusion
proteins of the present invention. Some of these methods
include: SDS-PAGE coupled with protein staining methods or
immunoblotting using anti-IgG, anti-HA and anti-G-CSF
antibodies. Other methods include matrix assisted laser
desporption/ionization-mass spectrometry (MALDI-MS), liquid
chromatography/mass spectrometry, isoelectric focusing,
analytical anion exchange, chromatofocussing, and circular
dichroism to name a few. A representative number of
heterologous fusion proteins were characterized using SDS-PAGE
coupled with immunoblotting as well as mass spectrometry

For example, Table 2 illustrates the calculated molecular
mass for a representative number of fusion proteins as well as
the observed mass (as measured by protease mapping/LC-MS).
The relative differences between observed mass and mass
calculated for a nonglycosylated protein are indicative of the
extent of glycosylation.

Table 2

Construct	Theoretical mass (kDa)	Observed mass (kDa)
C17A (GCSF)-IgG1 Fc	89.4	93.0
C17A (GCSF)-IgG4 Fc	88.7	92.7
C17A (GCSF)-HSA	86.0	85.1
C17A, A37N, S63N, E93N (GCSF)-IgG1 Fc	89.4	97-114 (104.7)
C17A, A37N, S63N, E93N (GCSF)-IgG4 Fc	88.7	96-112 (103.8)
C17A, A37N, S63N, E93N (GCSF)-HSA	86.0	87-98 (92.5)
C17A, A37N, W58N, Q67N (GCSF)-IgG1 Fc	89.3	95-110 (103.8)
C17A, A37N, W58N, Q67N (GCSF)-IgG4 Fc	88.8	95-110 (102.4)
C17A, A37N, W58N, Q67N (GCSF)-HSA	85.9	87-98 (93.1)

Administration and Preparation of Compositions:

The heterologous fusion proteins of the present invention
may be formulated with one or more excipients. The active

fusion proteins of the present invention may be combined with a pharmaceutically acceptable buffer, and the pH adjusted to provide acceptable stability, and a pH acceptable for adminstration such as parenteral administration.

5   Optionally, one or more pharmaceutically-acceptable anti-microbial agents may be added. Meta-cresol and phenol are preferred pharmaceutically-acceptable microbial agents. One or more pharmaceutically-acceptable salts may be added to adjust the ionic strength or tonicity. One or more excipients

10   may be added to adjust the isotonicity of the formulation. Glycerin is an example of an isotonicity-adjusting excipient. Pharmaceutically acceptable means suitable for administration to a human or other animal and thus, does not contain toxic elements or undersirable contaminants and does not interfere

15   with the activity of the active compounds therein.

A pharmaceutically-acceptable salt form of the heterologous fusion proteins of the present invention may be used in the present invention. Acids commonly employed to form acid addition salts are inorganic acids such as

20   hydrochloric acid, hydrobromic acid, hydriodic acid, sulfuric acid, phosphoric acid, and the like, and organic acids such as p-toluenesulfonic acid, methanesulfonic acid, oxalic acid, p-bromophenyl-sulfonic acid, carbonic acid, succinic acid, citric acid, benzoic acid, acetic acid, and the like.

25   Preferred acid addition salts are those formed with mineral acids such as hydrochloric acid and hydrobromic acid.

Base addition salts include those derived from inorganic bases, such as ammonium or alkali or alkaline earth metal hydroxides, carbonates, bicarbonates, and the like. Such

30   bases useful in preparing the salts of this invention thus include sodium hydroxide, potassium hydroxide, ammonium hydroxide, potassium carbonate, and the like.

Administration may be via any route known to be effective

35   by the physician of ordinary skill. Peripheral parenteral is one such method. Parenteral administration is commonly

understood in the medical literature as the injection of a
dosage form into the body by a sterile syringe or some other
mechanical device such as an infusion pump.   Peripheral
parenteral routes can include intravenous, intramuscular,
5   subcutaneous, and intraperitoneal routes of administration.

The heterologous fusion proteins of the present invention
may also be amenable to administration by oral, rectal, nasal,
or lower respiratory routes, which are non-parenteral routes.
Of these non-parenteral routes, the lower respiratory route
10   and the oral route are □referred.

The heterologous fusion proteins of the present
invention can be used to treat patients with insufficient
circulating neutrophil levels, typically those undergoing
cancer chemotherapy.

15   An "effective amount" of the heterologous fusion
protein is the quantity which results in a desired
therapeutic and/or prophylactic effect without causing
unacceptable side-effects when administered to a subject in
need of G-CSF receptor stimulation.   A "desired therapeutic
20   effect" includes one or more of the following: 1) an
amelioration of the symptom(s) associated with the disease
or condition; 2) a delay in the onset of symptoms associated
with the disease or condition; 3) increased longevity
compared with the absence of the treatment; and 4) greater
25   quality of life compared with the absence of the treatment.

The present invention comprises G-CSF compounds that
have improved biochemical and biophysical properties by
virtue of being fused to an albumin protein, an albumin
fragment, an albumin analog, a Fc protein, a Fc fragment, or
30   a Fc analog.   These heterologous proteins can be
successfully expressed in host cells, retain signaling
activities associated with activation of the G-CSF receptor,
and have prolonged half-lives.

EXAMPLES

Example 1: Construction of DNA encoding heterologous fusion proteins:

Table 1 provides the sequence of primers used to create functional glycosylation sites in different regions of the G-CSF protein (See Figure 1).

Table 1: Primer sequences used to introduce mutations into human G-CSF.

Mutation	A Primer*	B Primer*	C Primer*	D Primer*
WT	CF177[SEQ ID NO:26] *GTAAGCTTGCGT CGACGCTAGCGG CGCGCCGCC*ATG GCCGGACCTGCC ACCCAGAGCCCC ATGAAGCTG	CF178[SEQ ID NO:27] GGGGCAGGGAGC TGGCTGGGCCCA GTGGAGTGGCTT CCTGCACTGTCC AGAGTGCACTGT G	CF179[SEQ ID NO:28] GGACAGTGCAGG AAGCCACTCCAC TGGGCCCAGCCA GCTCCCTGCCCC AGAGCTTCCTG	CF176[SEQ ID NO:29] *GAACCTCGAGGA TCC*TCATTAGGG CTGGGCAAGGTG CCTTAAGACGCG GTACGACACCTC CAGGAAGCTCTG
C17A *SacI*	CF177[SEQ ID NO:30] *GTAAGCTTGCGT CGACGCTAGCGG CGCGCCGCC*ATG GCCGGACCTGCC ACCCAGAGCCCC ATGAAGCTG	C17Arev[SEQ ID NO:31] GCTCTAAGGCCT TGAGCAGGAAGC TCTGGGGCAGGG AGCTCGCTGGGC CCAGTGGAG	C17Afor[SEQ ID NO:32] GGGCCCAGCGAG CTCCCTGCCCCA GAGCTTCCTGCT CAAGGCCTTAGA GCAAG	CF176[SEQ ID NO:33] *GAACCTCGAGGA TCC*TCATTAGGG CTGGGCAAGGTG CCTTAAGACGCG GTACGACACCTC CAGGAAGCTCTG
A37N,Y39T *SpeI*	CF177[SEQ ID NO:34] *GTAAGCTTGCGT CGACGCTAGCGG CGCGCCGCC*ATG GCCGGACCTGCC ACCCAGAGCCCC ATGAAGCTG	A37Nrev[SEQ ID NO:35] GTCCGAGCAGCA CTAGTTCCTCGG GGTGGCACAGCT TGGTGGTGTTAC ACAGCTTCTCCT G	A37Nfor[SEQ ID NO:36] GGCGCAGCGCTC CAGGAGAAGCTG TGTAACACCACC AAGCTGTGCCAC CCCGAGGAACTA GTGCTG	CF176[SEQ ID NO:37] *GAACCTCGAGGA TCC*TCATTAGGG CTGGGCAAGGTG CCTTAAGACGCG GTACGACACCTC CAGGAAGCTCTG
T133N, G135T	CF177[SEQ ID NO:38]	T133Nrev[SEQ ID NO:39]	T133Nfor[SEQ ID NO:40]	CF176[SEQ ID NO:41]

Eco47III	GTAAGCTTGCGT CGACGCTAGCGG CGCGCCGCCATG GCCGGACCTGCC ACCCAGAGCCCC ATGAAGCTG	GCCCGGCGCTGG AAAGCGCTGGCG AAGGCCGGCATG GCGGTCTGGTTG GGCTGCAGGGCA G	GGCCCCTGCCCT GCAGCCCAACCA GACCGCCATGCC GGCCTTCGCCAG CGCTTTCCAGCG	GAACCTCGAGGA TCCTCATTAGGG CTGGGCAAGGTG CCTTAAGACGCG GTACGACACCTC CAGGAAGCTCTG
A141N, A143T SapI	CF177[SEQ ID NO:42] GTAAGCTTGCGT CGACGCTAGCGG CGCGCCGCCATG GCCGGACCTGCC ACCCAGAGCCCC ATGAAGCTG	A141Nrev[SEQ ID NO:43] GCCCGGCGCTGG AAGGTAGAGTTG AAGGCCGGCATG GCACCCTGGGTG GGCTGAAGAGCA GGGGCCAT	A141Nfor[SEQ ID NO:44] GGGAATGGCCCC TGCTCTTCAGCC CACCCAGGGTGC CATGCCGGCCTT CAACTCTACCTT CCAGCGCCGGGC AG	CF176[SEQ ID NO:45] GAACCTCGAGGA TCCTCATTAGGG CTGGGCAAGGTG CCTTAAGACGCG GTACGACACCTC CAGGAAGCTCTG

P57V, W58N, P60T HpaI	JCB128[SEQ ID NO:46] GCTAGCGGCGCG CCACCATG	JCB136[SEQ ID NO:47] GCTCAGGGTAGC GTTAACGATGCC CAGAGAGTG	JCB137[SEQ ID NO:48] GGGCATCGTTAA CGCTACCCTGAG CAGCTG	JCB129[SEQ ID NO:49] GACTCGAGGATC CTCATTAGGGCT GGG
Q67N, L69T NaeI	JCB134[SEQ ID NO:50] GCTAGCGGCGCG CCACCATGGCCG GACCTGCCACCC AG	JCB138[SEQ ID NO:51] CAAGCAGCCGGC CAGCTGGGTGGC GTTGCTGGGGCA GCTGCTCAG	JCB139[SEQ ID NO:52] GCCCCAGCAACG CCACCCAGCTGG CCGGCTGCTTGA G	JCB135[SEQ ID NO:53] GACTCGAGGATC CTCATTAGGGCT GGGCAAGGTGCC TTAAGACGCGG
P60N, S62T SpeI	JCB128[SEQ ID NO:54] GCTAGCGGCGCG CCACCATG	JCB130[SEQ ID NO:55] GGGGCAACTAGT CAGGTTAGCCCA GGG	JCB131[SEQ ID NO:56] GCTAACCTGACT AGTTGCCCCAGC CAG	JCB129[SEQ ID NO:57] GACTCGAGGATC CTCATTAGGGCT GGG
S63N, P65T MfeI	JCB128[SEQ ID NO:58] GCTAGCGGCGCG CCACCATG	JCB132[SEQ ID NO:59] GGTGCAATTGCT CAGGGGAGCCCA	JCB133[SEQ ID NO:60] GCAATTGCACCA GCCAGGCCCTG	JCB129[SEQ ID NO:61] GACTCGAGGATC CTCATTAGGGCT

		G		GGG
E93N,I95T *BspEI*	JCB134[SEQ ID NO:62] GCTAGCGGCGCG CCACCATGGCCG GACCTGCCACCC AG	JCB140[SEQ ID NO:63] CCGG**ACTGG**TCC C**GTT**CAGGGCCT GCAGGAGCCCCT G	JCB141[SEQ ID NO:64] GAACGGG**ACCAG** **TCC**GGAGTTGGG TCCCACCTTGG	JCB135[SEQ ID NO:65] ***GACT***CGAGGATC CTCATTAGGGCT GGGCAAGGTGCC TTAAGACGCGG

| *SalI* | JCB155[SEQ ID NO:66] ***GTCGAC***GCTAGC GGCGCGCCACCA TGGCCGGACCTG | | | |

*Nucleotides in bold represent changes imposed in the target sequence and nucleotides in bold <u>and</u> italics represent flanking sequences which may add restriction sites to facilitate cloning, Kozac sequences, or stop codons.

Preparation 1a: DNA encoding wild-type human G-CSF

A strand overlapping extension PCR reaction was used to create a wild type human G-CSF construct in order to eliminate the methylation of an ApaI site. Isolated human G-CSF cDNA served as the template for these reactions. The 5' end A primer was used to create a restriction enzyme site prior to the start of the coding region as well as to introduce a Kozac sequence (GGCGCC) 5' of the coding leader sequence to faciliate translation in cell culture.

The A-B product was generated using primers CF177 and CF178 in a PCR reaction. Likewise, the C-D product was produced with primers CF179 and CF176. The products were isolated and combined. The combined mixture was then used as a template with primers CF177 and CF178 to create the full-length wild-type construct. [Nelson, R.M. and Long, G.C. (1989), *Anal. Biochem.* 180:147-151].

The full-length product was ligated into the pCR2.1-Topo vector (Invitrogen, Inc. Cat. No. K4500-40) by way of a topoisomerase TA overhang system to create pCR2.1G-CSF.

The following protocol was used for preparation of the full-length wild-type G-CSF protein as well as each of the G-CSF analogs. Approximately 5 ng of template DNA and 15 pmol of each primer was used in the initial PCR reactions.

5 The reactions were prepared using Platinum PCR Supermix® (GibcoBRL Cat. No. 11306-016). The PCR reactions were denatured at 94^{0}C for 5 min and then subject to 25 cycles wherein each cycle consisted of 30 seconds at 94^{0}C followed by 30 seconds at 60^{0}C followed by 30 seconds at 72^{0}C. A

10 final extension was carried out for 7 minutes at 72^{0}C. PCR fragments were isolated from agarose gels and purified using a Qiaquick® gel extraction kit (Qiagen, Cat. No. #28706). DNA was resuspended in sterile water and used for the final PCR reaction to prepare full-length product.

15

Preparation 1b: DNA encoding G-CSF[A37N,Y39T, P57V,W58N,P60T,Q67N,L69T] was constructed as follows:

DNA encoding G-CSF[A37N,Y39T,Q67N,L69T] was subcloned into pJB02 to create pJB02G-CSF[A37N,Y39T,Q67N,L69T] and

20 pJB02G-CSF[A37N,Y39T,P57V,W58N,P60T] served as the template for strand overlapping expression PCR. JCB155 and JCB136 served as the A and B primers and JCB137 and JCB135 served as the C and D primers. The full-length mutated cDNA was prepared as described previously using JCB155 and JCB134

25 primers. The resulting full-length DNA encodes a protein with consensus N-linked glycosylation sites in region 1, region 2, and region 9 of the protein (See Figure 1). The full-length cDNA was ligated back into pCR2.1-Topo to create pCR2.1G-CSF[A37N,Y39T,P57V,W58N,P60T,

30 Q67N,L69T].

Preparation 1c: DNA encoding G-CSF[A37N,Y39T, S63N,P64T,E93N,I95T] was constructed as follows:

DNA encoding G-CSF[A37N,Y39T,E93N,I95T] was subcloned

35 into pJB02 to create pJB02G-CSF[A37N,Y39T,E93N,I95T] and pJB02G-CSF[A37N,Y39T,E93N,I95T] served as the template for

strand overlapping expression PCR. JCB155 and JCB132 served
as the A and B primers and JCB133 and JCB135 served as the C
and D primers. The full-length mutated cDNA was prepared as
described previously using JCB155 and JCB135 primers. The

5   resulting full-length DNA encodes a protein with consensus
N-linked glycosylation sites in region 1, region 7, and
region 10 of the protein (See Figure 1). The full-length
cDNA was ligated back into pCR2.1-Topo to create pCR2.1G-
CSF[A37N,Y39T,S63N,P64T,E93N,I95T].

10

Preparation 1d: DNA encoding G-CSF[C17A] which is G-CSF
wherein the amino acid at position 17 is substituted with
Ala is constructed as follows:

    The wild-type construct in the pCR2.1-Topo vector

15 (pCR2.1G-CSF) serves as the PCR template for the C17A
mutatgenesis. Strand ovelapping extension PCR is performed
as described previously. CF177 and C17Arev serve as the A-B
primers and C17Afor and CF176 serve as the C-D primers. The
full-length mutated cDNA is prepared as described previously

20 using the CF177 and CF176 primers. The B and C primers are
used to mutate the DNA such that a SacI restriction site is
created and the protein expressed from the full-length
sequence contains an Alanine instead of a Cysteine at
position 17. The full-length cDNA is ligated back into the

25 pCR2.1-Topo vector to create pCR2.1G-CSF[C17A] wherein the
sequence is confirmed. G-CSF analog encoding DNA is then
cloned into the Nhe/Xho sites of mammalian expression vector
pJB02 to create pJB02G-CSF[C17A].

30   Preparation 1e: DNA encoding G-CSF[A37N,Y39T] is
constructed as follows:
    Strand overlapping extension PCR is performed using
pCR2.1G-CSF[C17A] as the template. Primers CF177 and
A37Nrev serve as the A-B primers and CF176 and A37Nfor serve

35 as the C-D primers. The full-length mutated cDNA is prepared
as described previously using the CF177 and CF176 primers.

The B and C primers contain mismatched sequences such that a
SpeI site is created in the DNA and the protein expressed
from the full-length sequence contains a consensus sequence
for N-linked glycosylation in region 1 of the protein. The
5   full-length cDNA is ligated back into the pCR2.1-Topo vector
to create pCR2.1G-CSF[A37N,Y39T] wherein the sequence is
confirmed. G-CSF analog encoding DNA is then cloned into the
Nhe/Xho sites of mammalian expression vector pJB02 to create
pJB02G-CSF[A37N,Y39T].

10

Preparation 1f: DNA encoding G-CSF[P57V,W58N,P60T] is
constructed as follows:
Strand overlapping extension PCR is performed using
pJB02G-CSF[C17A] as the template. Primers JCB128 and JCB136
15   serve as the A-B primers and JCB137 and JCB129 serve as the
C-D primers. The full-length mutated cDNA is prepared as
described previously using the JCB128 and JCB129 primers.
The B and C primers contain mismatched sequences such that a
HpaI site is created and the protein expressed from the
20   full-length sequence contains a consensus sequence for N-
linked glycosylation in region 2 of the protein. The full-
length cDNA is ligated back into the pCR2.1-Topo vector to
create pCR2.1G-CSF[P57V,W58N,P60T] wherein the sequence is
confirmed. G-CSF analog encoding DNA is then cloned into the
25   *Nhe/Xho* sites of mammalian expression vector pJB02 to create
pJB02G-CSF[P57V,W58N,P60T].

Preparation 1g: DNA encoding G-CSF[P60N,S62T] is
constructed as follows:
30   Strand overlapping extension PCR is performed using
pJB02G-CSF[C17A] as the template. Primers JCB128 and JCB130
serve as the A-B primers and JCB131 and JCB129 serve as the
C-D primers. The full-length mutated cDNA is prepared as
described previously using the JCB128 and JCB129 primers.
35   The B and C primers contain mismatched sequences such that a
SpeI site is created and the protein expressed from the

full-length sequence contains a consensus sequence for N-linked glycosylation in region 4 of the protein. The full-length cDNA is ligated back into the pCR2.1-Topo vector to create pCR2.1G-CSF[P60N,S62T] wherein the sequence is
5 confirmed. G-CSF analog encoding DNA is then cloned into the *Nhe/Xho* sites of mammalian expression vector pJB02 to create pJB02G-CSF[P60N,S62T].

Preparation 1h: DNA encoding G-CSF[S63N,P65T] is
10 constructed as follows:
Strand overlapping extension PCR is performed using pJB02G-CSF[C17A] as the template. Primers JCB128 and JCB132 serve as the A-B primers and JCB133 and JCB129 serve as the C-D primers. The full-length mutated cDNA is prepared as
15 described previously using the JCB128 and JCB129 primers. The B and C primers contain mismatched sequences such that a *MfeI* site is created and the protein expressed from the full-length sequence contains a consensus sequence for N-linked glycosylation in region 7 of the protein. The full-
20 length cDNA is ligated back into the pCR2.1-Topo vector to create pCR2.1G-CSF[S63N,P65T] wherein the sequence is confirmed. G-CSF analog encoding DNA is then cloned into the *Nhe/Xho* sites of mammalian expression vector pJB02 to create pJB02G-CSF[S63N,P65T].

25
Preparation 1i: DNA encoding G-CSF[Q67N,L69T] is constructed as follows:
Strand overlapping extension PCR is performed using pJB02G-CSF[C17A] as the template. Primers JCB134 and JCB138
30 serve as the A-B primers and JCB139 and JCB135 serve as the C-D primers. The full-length mutated cDNA is prepared as described previously using the JCB128 and JCB129 primers. The B and C primers contain mismatched sequences such that a *NaeI* site is created and the protein expressed from the
35 full-length sequence contains a consensus sequence for N-linked glycosylation in region 9 of the protein. The full-

length cDNA is ligated back into the pCR2.1-Topo vector to create pCR2.1G-CSF[Q67N,L69T] wherein the sequence is confirmed. G-CSF analog encoding DNA is then cloned into the *Nhe/Xho* sites of mammalian expression vector pJB02 to create
5   pJB02G-CSF[Q67N,L69T].

Preparation 1j: DNA encoding G-CSF[E93N,I95T] is constructed as follows:
Strand overlapping extension PCR is performed using
10   pJB02G-CSF[C17A] as the template. Primers JCB134 and JCB140 serve as the A-B primers and JCB141 and JCB135 serve as the C-D primers. The full-length mutated cDNA is prepared as described previously using the JCB128 and JCB129 primers. The B and C primers contain mismatched sequences such that
15   a *BspEI* site is created and the protein expressed from the full-length sequence contains a consensus sequence for N-linked glycosylation in region 10 of the protein. The full-length cDNA is ligated back into the pCR2.1-Topo vector to create pCR2.1G-CSF[E93N,I95T] wherein the sequence is
20   confirmed. G-CSF analog encoding DNA is then cloned into the *Nhe/Xho* sites of mammalian expression vector pJB02 to create pJB02G-CSF[E93T,I95T].

Preparation 1k: DNA encoding G-CSF[T133N,G135T] is
25   constructed as follows:
Strand overlapping extension PCR is performed using pCR2.1G-CSF[C17A] as the template. Primers CF177 and T133Nrev serve as the A-B primers and T133Nfor and CF176 serve as the C-D primers. The full-length mutated cDNA is
30   prepared as described previously using the CF177 and CF176 primers. The B and C primers contain mismatched sequences such that an *Eco47III* site is created and the protein expressed from the full-length sequence contains a consensus sequence for N-linked glycosylation in region 13 of the
35   protein. The full-length cDNA is ligated back into the

pCR2.1-Topo vector to create pCR2.1G-CSF[T133N,G135T] wherein the sequence is confirmed.

Preparation 11: DNA encoding G-CSF[A141N,A143T] is
5   constructed as follows:

Strand overlapping extension PCR is performed using pCR2.1G-CSF[C17A] as the template. Primers CF177 and A141Nrev serve as the A-B primers and A141Nfor and CF176 serve as the C-D primers. The full-length mutated cDNA is
10 prepared as described previously using the CF177 and CF176 primers. The B and C primers contain mismatched sequences such that an SapI site is created and the protein expressed from the full-length sequence contains a consensus sequence for N-linked glycosylation in region 14 of the protein. The
15 full-length cDNA is ligated back into the pCR2.1-Topo vector to create pCR2.1G-CSF[A141N,A143T] wherein the sequence is confirmed.

Preparation 1m: DNA encoding G-CSF[A37N,Y39T,T133N,
20 G135T] is constructed as follows:

A 210 bp insert containing G-CSF[A37N,Y39T] is isolated from pCR2.1G-CSF[A37N,Y39T] using EcoNI. This fragment is ligated into pCR2.1G-CSF[T133N,G135T] which is prepared by cleavage with EcoNI and subsequent isolation of the vector
25 (4359 bp) from a 210 bp fragment containing wild-type G-CSF sequences. This ligation creates pCR2.1G-CSF[A37N,Y39T,T133N,G135T]. Analog encoding DNA is then subcloned into pJB02 using NheI/XhoI to create pJB02G-CSF[A37N,Y39T,T133N,G135T].

30

Preparation 1n: DNA encoding G-CSF[A37N,Y39T,A141N,
A143T] is constructed as follows:

A 210 bp insert containing G-CSF[A37N,Y39T] is isolated from pCR2.1G-CSF[A37N,Y39T] using EcoNI. This fragment is
35 ligated into pCR2.1G-CSF[A141N,A143T] which is prepared by cleavage with EcoNI and subsequent isolation of the vector

(4359 bp) from a 210 bp fragment containing wild-type G-CSF sequences. This ligation creates pCR2.1G-CSF[A37N,Y39T,A141N,A143T]. Analog encoding DNA is then subcloned into pJB02 using *NheI/XhoI* to create pJB02G-
5 CSF[A37N,Y39T,A141N,A143T].

Preparation 1o: DNA encoding G-CSF[A37N,Y39T,P57V,W58N,P60T] is constructed as follows:

DNA encoding G-CSF[A37N,Y39T] is subcloned into pJB02
10 to create pJB02G-CSF[A37N,Y39T] and pJB02G-CSF[A37N,Y39T] serves as the template for strand overlapping expression PCR. JCB128 and JCB136 serve as the A and B primers and JCB137 and JCB129 serve as the C and D primers. The full-length mutated cDNA is prepared as described previously
15 using JCB128 and JCB129 primers. The resulting full-length DNA encodes a protein with consensus N-linked glycosylation sites in region 1 and region 2 of the protein. The full-length cDNA is ligated back into pCR2.1-Topo to create pCR2.1G-CSF[A37N,Y39T,P57V,W58N,P60T].
20

Preparation 1p: DNA encoding G-CSF[A37N,Y39T,Q67N,L69T] is constructed as follows:

DNA encoding G-CSF[A37N,Y39T] is subcloned into pJB02 to create pJB02G-CSF[A37N,Y39T] and pJB02G-CSF[A37N,Y39T]
25 serves as the template for strand overlapping expression PCR. JCB134 and JCB138 serve as the A and B primers and JCB139 and JCB135 serve as the C and D primers. The full-length mutated cDNA is prepared as described previously using JCB128 and JCB129 primers. The resulting full-length
30 DNA encodes a protein with consensus N-linked glycosylation sites in region 1 and region 9 of the protein. The full-length cDNA is ligated back into pCR2.1-Topo to create pCR2.1G-CSF[A37N,Y39T,Q67N,L69T].

35 Preparation 1q: DNA encoding G-CSF[A37N,Y39T,E93N,I95T] is constructed as follows:

DNA encoding G-CSF[A37N,Y39T] is subcloned into pJB02 to
create pJB02G-CSF[A37N,Y39T] and pJB02G-CSF[A37N,Y39T]
serves as the template for strand overlapping expression
PCR. JCB134 and JCB140 serve as the A and B primers and
5   JCB141 and JCB135 serve as the C and D primers. The full-
length mutated cDNA is prepared as described previously
using JCB128 and JCB129 primers. The resulting full-length
DNA encodes a protein with consensus N-linked glycosylation
sites in region 1 and region 10 of the protein. The full-
10   length cDNA is ligated back into pCR2.1-Topo to create
pCR2.1G-CSF[A37N,Y39T,E93N,I95T].

     Various mutated G-CSF polypeptides were then ligated
in-frame directly to DNA encoding IgG1 or IgG4. Various
15   mutated G-CSF polypeptides were also ligated in-frame via
DNA encoding a peptide linker or directly to human albumin.

     Example 2: Expression of heterologus fusion proteins:

20   2a: Expression in 293/EBNA cells:
     Each full-length DNA encoding a heterologous fusion
protein was cloned into the pJB02. This vector contains
both the Ori P and Epstein Barr virus nuclear antigen (EBNA)
components which are necessary for sustained, transient
25   expression in 293 EBNA cells. Further, this expression
plasmid contains a puromycin resistance gene expressed from
the CMV promoter as well as an ampicillin resistance gene.
The gene of interest is also expressed from the CMV
promoter.
30   The transfection mixture was prepared by mixing 73 μl
of the liposome transfection agent Fugene 6® (Roche
Molecular Biochemicals, Cat. No. 1815-075) with 820 μl Opti-
Mem® (GibcoBRL Cat. No. 31985-062). G-CSF pJB02 DNA
(12μg), prepared using a Qiagen plasmid maxiprep kit
35   (Qiagen, Cat. No. 12163), was then added to the mixture.

The mixture was incubated at room temperature for 15 minutes.

Cells were plated on 10 cm^2 plates in DMEM/F12 3:1 (GibcoBRL Cat. No. 93-0152DK) supplemented with 5% fetal
5   bovine serum, 20mM HEPES, 2 mM L-glutamine, and 50 µg/mL Geneticin such that the plates were 60% to 80% confluent by the time of the transfection. Immediately before the transfection mixture was added to the plates, fresh media was added. The mixture was then added dropwise to cells
10   with intermittent swirling. Plates were then incubated at 37°C in a 5% CO$_2$ atmosphere for 24 hours at which point the media was changed to hybritech medium without serum. The media containing a secreted form of a glycosylated G-CSF analog was then isolated 48 hours later.
15

2b: Expression in CHO cells:

The expression vector for expression in CHO-K1 cells is pEE14.1. This vector includes the glutamine synthetase gene which enables selection using methionine sulfoximine. This
20   gene includes two poly A signals at the 3' end. G-CSF analogs are expressed from the CMV promoter which includes 5' untranslated sequences from the hCMV-MIE gene to enhance mRNA levels and translatability. The SV40 poly A signal is cloned 3' of the fusion protein DNA. The SV40 late promoter
25   drives expression of GS minigene. This expression vector encoding the gene of interest is prepared for transfection using a QIAGEN Maxi Prep Kit (QIAGEN, Cat. No. 12362). The final DNA pellet (50-100 µg) is resuspended in 100 µl of basal formulation medium (GibcoBRL CD-CHO Medium without L-
30   Glutamine, without thymidine, without hypoxanthine).

Before each transfection, CHO-K1 cells are counted and checked for viability. A volume equal to 1 x 10^7 cells is centrifuged and the cell pellet rinsed with basal formulation medium. The cells are centrifuged a second time
35   and the final pellet resuspended in basal formulation medium (700 µl final volume).

The resuspended DNA and cells are then mixed together in a standard electroporation cuvette (Gene Pulsar Cuvette) used to support mammalian transfections, and placed on ice for five minutes. The cell/DNA mix is then electroporated
5 in a BioRad Gene Pulsar device set at 300V/975 µF and the cuvette placed back on ice for five minutes. The cell/DNA mixed is diluted into 20 ml of cell growth medium in a non-tissue culture treated T75 flask and incubated at 37°C / 5% CO_2 for 48-72 hours.

10 The cells are counted, checked for viability, and plated at various cell densities in selective medium in 96 well tissue culture plates and incubated at 37°C in a 5% CO_2 atmosphere. Selective medium is basal medium with 1X HT Supplement (GibcoBRL 100X HT Stock), 100 µg/mL Dextran
15 Sulfate (Sigma 100 mg/ml stock), 1X GS Supplements (JRH BioSciences 50X Stock) and 25 µM MSX (Methionine Sulphoximine). The plates are monitored for colony formation and screened for glycosylated G-CSF analog production.

20

Example 3: Purification of Heterologous Fusion Proteins

Heterologous fusion proteins containing a hyperglycosylated G-CSF protein fused to albumin were
25 harvested from host cells and then dialyzed against 20 mM Tris pH 7.4. An anion exchange column (1 ml Pharmacia HiTrap Q) was equilibrated with 20 mM Tris pH 7.4 and the dialyzed material loaded at 2 ml/min. The protein was eluted from the column using a linear gradient from 0 to 500
30 mM NaCl in 80 min at 1 ml/min and elution was monitored by UV absorbance at 280 mm. SDS-PAGE analysis was used to identify and pool fractions of interest. This pool was dialyzed against 25 mM sodium acetate (NaOAc) pH 5.0

A cation exchange column (1 ml Pharmacia HiTrap S
35 column) was equilibrated with 25 mM NaOAc pH 5.0 and the dialysate was loaded at 1 ml/min. The protein was eluted

from the column using a linear gradient from 0 to 500 mM
NaCl in 30 min.   The fractions were immediately neutralized
with 1 M Tris pH 8 to a final pH of 7.   SDS-PAGE gels were
used to identify and pool fractions of interest.

5   Heterologous fusion proteins containing a
hyperglycosylated G-CSF protein fused to an Fc protein were
harvested from host cells and dialyzed against 20 mM sodium
phosphate pH 7.0.   An affinity column (1 ml Pharmacia HiTrap
Protein A or rProtein A) was equilibrated with 20 mM sodium

10   phosphate pH 7.0 and the dialysate was loaded at 2 ml/min.
1 ml/min of 100 mM citric acid pH 3 was used to elute the
protein.   Fractions were immediately neutralized with 1M
Tris pH 8 to pH 7 and peak fractions (determined by in-line
OD280 monitoring) were further diluted with 20 mM sodium

15   phosphate pH 7.0. .   SDS-PAGE analysis was used to identify
and pool fractions of interest.

WE CLAIM:

1. A heterologous fusion protein comprising a
5 hyperglycosylated G-CSF analog fused to a polypeptide
selected from the group consisting of
      a) human albumin;
      b) human albumin analogs; and
      c) fragments of human albumin.
10

2. The heterologous fusion protein of claim 1, wherein the
hyperglycosylated G-CSF analog is fused to the polypeptide
via a peptide linker.

15 3. The heterologous fusion protein of Claim 2 wherein the
peptide linker is selected from the group consisting of:
      a) a glycine rich peptide;
      b) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]$_n$
         where n is 1, 2, 3, 4, or 5; and
20   c) a peptide having the sequence [Gly-Gly-Gly-Gly-
         Ser]$_3$.

4. The heterologous fusion protein of Claims 1, 2, or 3
wherein the hyperglycosylated G-CSF analog comprises the
25 amino acid sequence of the formula I: [SEQ ID NO: 1]

```
1                  5                  10                 15
Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
            20                 25                 30
Xaa Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
        35                 40                 45
Glu Lys Leu Cys Xaa Xaa Xaa Lys Leu Cys His Pro Glu Glu Leu Val
        50                 55                 60
Leu Leu Gly His Ser Leu Gly Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
    65                 70                 75                 80
Xaa Xaa Xaa Xaa Xaa Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
                85                 90                 95
Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Xaa Xaa Xaa Ser
            100                105                110
Xaa Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
        115                120                125
Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
        130                135                140
Ala Leu Gln Pro Xaa Xaa Xaa Ala Met Pro Ala Phe Xaa Xaa Xaa Phe
    145                150                155                160
```

```
Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Aln Ser Phe
                165                     170
Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro       (I)
```

5   wherein:
        Xaa at position 17 is Cys, Ala, Leu, Ser, or Glu;
        Xaa at position 37 is Ala or Asn;
        Xaa at position 38 is Thr, or any other amino acid except
        Pro;
10   Xaa at position 39 is Tyr, Thr, or Ser;
        Xaa at position 57 is Pro or Val;
        Xaa at position 58 is Trp or Asn;
        Xaa at position 59 is Ala or any other amino acid except
        Pro;
15   Xaa at position 60 is Pro, Thr, Asn, or Ser,
        Xaa at position 61 is Leu, or any other amino acid except
        Pro;
        Xaa at position 62 is Ser or Thr;
        Xaa at position 63 is Ser or Asn;
20   Xaa at position 64 is Cys or any other amino acid except
        Pro;
        Xaa at position 65 is Pro, Ser, or Thr;
        Xaa at position 66 is Ser or Thr;
        Xaa at position 67 is Gln or Asn;
25   Xaa at position 68 is Ala or any other amino acid except
        Pro;
        Xaa at position 69 is Leu, Thr, or Ser
        Xaa at position 93 is Glu or Asn
        Xaa at position 94 is Gly or any other amino acid except
30   Pro;
        Xaa at position 95 is Ile, Asn, Ser, or Thr;
        Xaa at position 97 is Pro, Ser, Thr, or Asn;
        Xaa at position 133 is Thr or Asn;
        Xaa at position 134 is Gln or any other amino acid except
35   Pro;
        Xaa at position 135 is Gly, Ser, or Thr
        Xaa at position 141 is Ala or Asn;
        Xaa at position 142 is Ser or any other amino acid except
        Pro; and
40   Xaa at position 143 is Ala, Ser, or Thr;

        and wherein:
        Xaa at positions 37, 38, and 39 constitute region 1;
        Xaa at positions 58, 59, and 60 constitute region 2;
45   Xaa at positions 59, 60, and 61 constitute region 3;
        Xaa at positions 60, 61, and 62 constitute region 4;

Xaa at positions 61, 62, and 63 constitute region 5;

Xaa at positions 62, 63, and 64 constitute region 6;

Xaa at positions 63, 64, and 65 constitute region 7;

Xaa at positions 64, 65, and 66 constitute region 8;

5 Xaa at positions 67, 68, and 69 constitute region 9;

Xaa at positions 93, 94, and 95 constitute region 10;

Xaa at positions 94, 95, and Ser at position 96
constitute region 11;

Xaa at positions 95, and 97, and Ser at position 96
10 constitute region 12;

Xaa at positions 133, 134, and 135 constitute
region 13;

Xaa at positions 141, 142, and 143 constitute
region 14;

15

and provided that at least one of regions 1 through 14
comprises the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any
amino acid except Pro and Xaa2 is Ser or Thr.

20 5. The heterologous fusion protein of Claim 4 wherein any
two regions of regions 1 through 14 comprise the sequence
Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and
Xaa2 is Ser or Thr.

25 6. The heterologous fusion protein of Claim 4 wherein any
three regions of regions 1 through 14 comprise the sequence
Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and
Xaa2 is Ser or Thr.

30 7. The heterologous fusion protein of Claim 4 wherein any
four regions of regions 1 through 14 comprise the sequence
Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and
Xaa2 is Ser or Thr

8. The heterologous fusion protein of Claim 4 wherein the hyperglycosylated G-CSF analog is selected from the group consisting of:

a) G-CSF[A37N,Y39T]

5 b) G-CSF[P57V,W58N,P60T]

c) G-CSF[P60N,S62T]

d) G-CSF[S63N,P65T]

e) G-CSF[Q67N,L69T]

f) G-CSF[E93N,I95T]

10 g) G-CSF[T133N,G135T]

h) G-CSF[A141N,A143T]

i) G-CSF[A37N,Y39T,P57V,W58N,P60T]

j) G-CSF[A37N,Y39T,P60N,S62T]

k) G-CSF[A37N,Y39T,S63N,P65T]

15 l) G-CSF[A37N,Y39T,Q67N,L69T]

m) G-CSF[A37N,Y39T,E93N,I95T]

n) G-CSF[A37N,Y39T,T133N,G135T]

o) G-CSF[A37N,Y39T,A141N,A143T]

p) G-CSF[A37N,Y39T,P57V,W58N,P60T,S63N,P65T]

20 q) G-CSF[A37N,Y39T,P57V,W58N,P60T,Q67N,L69T]

r) G-CSF[A37N,Y39T,S63N,P65T,E93N,I95T]

9. The heterologous fusion protein of claim 8, wherein the hyperglycosylated G-CSF analog is G-CSF[A37N,

25 Y39T,P57V,W58N,P60T,Q67N,L69T].

10. The heterologous fusion protein of claim 8, wherein the hyperglycosylated G-CSF analog is G-CSF[A37N,Y39T,S63N,P65T,E93N,I95T].

30

11. A heterologous fusion protein which is the product of the expression in a host cell of an exogenous DNA sequence which comprises a DNA sequence encoding a heterologous fusion protein of any one of Claims 1 through 10.

35

12. A polynucleotide encoding a heterologous fusion protein of any one of Claims 1 through 11.

13. A polynucleotide which comprises a DNA sequence
5   selected from the group consisting of:
a) SEQ ID NO:2
b) SEQ ID NO:3
c) SEQ ID NO:4
d) SEQ ID NO:5
10 e) SEQ ID NO:6
f) SEQ ID NO:7
g) SEQ ID NO:8
h) SEQ ID NO:9
i) SEQ ID NO:10
15 j) SEQ ID NO:11
k) SEQ ID NO:12
l) SEQ ID NO:13
m). SEQ ID NO:14
n) SEQ ID NO:15
20 o) SEQ ID NO:16,
fused in-frame to a DNA encoding a protein selected from the group consisting of:
         a) human albumin,
         b) human albumin analog; and
25   c) fragments of human albumin.

14. The polynucleotide of Claim 13, wherein the DNA fused in-frame comprises SEQ ID NO: 17.

30 15. The heterologous fusion protein of any one of Claims 1 through 11 wherein the polypeptide is human albumin.

16. The heterologous fusion protein of any one of Claims 1 through 11 wherein the polypeptide is an N-terminal fragment
35 of albumin.

17. A method for increasing neutrophil levels in a mammal comprising administering a therapeutically effective amount of the heterologous fusion protein of any one of Claims 1
40 through 11, 15 and 16.

18. The use of the heterologous fusion protein as claimed
in any one of Claims 1 through 11, 15 and 16 for the
manufacture of a medicament for the treatment of patients
5   with insufficient circulating neutrophil levels.

19. Use of a heterologous fusion protein of any one of
Claims 1 through 11, 15, and 16 as a medicament.

10   20. Use of a heterologous fusion protein of any one of
Claims 1 through 11, 15, and 16 for the treatment of
patients with insufficient circulating neutrophil levels.

21. A pharmaceutical formulation adapted for the treatment
15   of patients with insufficient neutrophil levels comprising a
heterologous fusion protein of any one of Claims 1 through
11, 15, and 16.

22. A heterologous fusion protein comprising a
20   hyperglycosylated G-CSF analog fused to a polypeptide
selected from the group consisting of:
        a) the Fc portion of an immunoglobulin;
        b) an analog of the Fc portion of an immunoglobulin;
            and
25   c) fragments of the Fc portion of an immunoglobulin.

23. The heterologous fusion protein of Claim 22, wherein
the hyperglycosylated G-CSF analog is fused to the
polypeptide via a peptide linker.
30
24. The heterologous fusion protein of the Claim 23 wherein
the peptide linker is selected from the group consisting of:
        a) a glycine rich peptide;
        b) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]$_n$
35   where n is 1, 2, 3, 4, or 5; and
        c) a peptide having the sequence [Gly-Gly-Gly-Gly-Ser]$_3$.

25. The heterologous fusion protein of Claims 22, 23 or 24, wherein the hyperglycosylated G-CSF analog comprises the amino acid sequence of the formula I: [SEQ ID NO: 1]

```
    1                   5                   10                  15
    Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
                20                      25                  30
    Xaa Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
                35                      40                  45
    Glu Lys Leu Cys Xaa Xaa Xaa Lys Leu Cys His Pro Glu Glu Leu Val
                50                      55                  60
    Leu Leu Gly His Ser Leu Gly Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
    65                      70                  75                  80
    Xaa Xaa Xaa Xaa Xaa Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
                        85                  90                  95
    Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Xaa Xaa Xaa Ser
                    100                     105                 110
    Xaa Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
                115                     120                 125
    Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
                130                     135                 140
    Ala Leu Gln Pro Xaa Xaa Xaa Ala Met Pro Ala Phe Xaa Xaa Xaa Phe
    145                     150                     155                 160
    Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Aln Ser Phe
                    165                     170
    Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro        (I)
```

wherein:
Xaa at position 17 is Cys, Ala, Leu, Ser, or Glu;
Xaa at position 37 is Ala or Asn;
Xaa at position 38 is Thr, or any other amino acid except Pro;
Xaa at position 39 is Tyr, Thr, or Ser;
Xaa at position 57 is Pro or Val;
Xaa at position 58 is Trp or Asn;
Xaa at position 59 is Ala or any other amino acid except Pro;
Xaa at position 60 is Pro, Thr, Asn, or Ser,
Xaa at position 61 is Leu, or any other amino acid except Pro;
Xaa at position 62 is Ser or Thr;
Xaa at position 63 is Ser or Asn;
Xaa at position 64 is Cys or any other amino acid except Pro;
Xaa at position 65 is Pro, Ser, or Thr;
Xaa at position 66 is Ser or Thr;
Xaa at position 67 is Gln or Asn;
Xaa at position 68 is Ala or any other amino acid except Pro;
Xaa at position 69 is Leu, Thr, or Ser

Xaa at position 93 is Glu or Asn
Xaa at position 94 is Gly or any other amino acid except
Pro;
Xaa at position 95 is Ile, Asn, Ser, or Thr;
5   Xaa at position 97 is Pro, Ser, Thr, or Asn;
Xaa at position 133 is Thr or Asn;
Xaa at position 134 is Gln or any other amino acid except
Pro;
Xaa at position 135 is Gly, Ser, or Thr
10 Xaa at position 141 is Ala or Asn;
Xaa at position 142 is Ser or any other amino acid except
Pro; and
Xaa at position 143 is Ala, Ser, or Thr;

15   and wherein:
           Xaa at positions 37, 38, and 39 constitute region 1;
           Xaa at positions 58, 59, and 60 constitute region 2;
           Xaa at positions 59, 60, and 61 constitute region 3;
           Xaa at positions 60, 61, and 62 constitute region 4;
20   Xaa at positions 61, 62, and 63 constitute region 5;
           Xaa at positions 62, 63, and 64 constitute region 6;
           Xaa at positions 63, 64, and 65 constitute region 7;
           Xaa at positions 64, 65, and 66 constitute region 8;
           Xaa at positions 67, 68, and 69 constitute region 9;
25   Xaa at positions 93, 94, and 95 constitute region 10;
           Xaa at positions 94, 95, and Ser at position 96
                 constitute region 11;
           Xaa at positions 95, and 97, and Ser at position 96
                 constitute region 12;
30   Xaa at positions 133, 134, and 135 constitute
                 region 13;
           Xaa at positions 141, 142, and 143 constitute
                 region 14;

35   and provided that at least one of regions 1 through 14
comprises the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any
amino acid except Pro and Xaa2 is Ser or Thr.

26. The heterologous fusion protein of Claim 25 wherein any two regions of regions 1 through 14 comprise the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and 5 Xaa2 is Ser or Thr.

27. The heterologous fusion protein of Claim 25 wherein any three regions of regions 1 through 14 comprise the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and 10 Xaa2 is Ser or Thr.

28. The heterologous fusion protein of Claim 25 wherein any four regions of regions 1 through 14 comprise the sequence Asn Xaa1 Xaa2 wherein Xaa1 is any amino acid except Pro and 15 Xaa2 is Ser or Thr

29. The heterologous fusion protein of Claim 25 wherein the hyperglycosylated G-CSF analog is selected from the group consisting of:
20   a) G-CSF[A37N,Y39T]
  b) G-CSF[P57V,W58N,P60T]
  c) G-CSF[P60N,S62T]
  d) G-CSF[S63N,P65T]
  e) G-CSF[Q67N,L69T]
25   f) G-CSF[E93N,I95T]
  g) G-CSF[T133N,G135T]
  h) G-CSF[A141N,A143T]
  i) G-CSF[A37N,Y39T,P57V,W58N,P60T]
  j) G-CSF[A37N,Y39T,P60N,S62T]
30   k) G-CSF[A37N,Y39T,S63N,P65T]
  l) G-CSF[A37N,Y39T,Q67N,L69T]
  m) G-CSF[A37N,Y39T,E93N,I95T]
  n) G-CSF[A37N,Y39T,T133N,G135T]
  o) G-CSF[A37N,Y39T,A141N,A143T]
35   p) G-CSF[A37N,Y39T,P57V,W58N,P60T,S63N,P65T]
  q) G-CSF[A37N,Y39T,P57V,W58N,P60T,Q67N,L69T]

r) G-CSF[A37N,Y39T,S63N,P65T,E93N,I95T]

30.   The heterologous fusion protein of Claim 29 wherein the hyperglycosylated G-CSF analog is G-
CSF[A37N,Y39T,P57V,W58N,P60T,Q67N,L69T]

31.   The heterologous fusion protein of Claim 29 wherein the hyperglycosylated G-CSF analog is G-
CSF[A37N,Y39T,S63N,P65T,E93N,I95T]

32.   The heterologous fusion protein of any one of Claims 22 through 31 wherein the polypeptide is the Fc portion of an Ig selected from the group consisting of: IgG1, IgG2, IgG3, IgG4, IgE, IgA, IgD, or IgM.

33.   The heterologous fusion protein of Claim 32 wherein the polypeptide is the Fc portion of an Ig selected from the group consisting of: IgG1, IgG2, IgG3, and IgG4.

34.   The heterologous fusion protein of Claim 33 wherein the polypeptide is the Fc portion of an IgG1 immunoglobulin.

35.   The heterologous fusion protein of Claim 33 wherein the polypeptide is the Fc portion of an IgG4 immunoglobulin.

36.   The heterologous fusion protein of any one of Claims 22 through 35 wherein the Fc portion is a human IgG protein.

37.   The heterologous fusion protein of any one of Claims 22 through 36 wherein the Fc portion comprises hinge, CH2, and CH3 domains.

38.   The heterologous fusion protein of Claim 34 wherein the polypeptide has the sequence of SEQ ID NO: 33.

39. A polynucleotide encoding a heterologous fusion protein of Claim 34, wherein the polynucleotide comprises SEQ ID NO: 22.

5 40. A polynucleotide encoding a heterologous fusion protein of any one of Claims 22 through 38.

41. A vector comprising the polynucleotide of Claim 12, 13, 14, 39 and 40.

10

42. A host cell comprising the vector of Claim 41.

43. A host cell expressing at least one heterologous fusion protein of any one of Claims 1 through 11, 15, 16, and 22

15 through 38.

44. The host cell of Claim 43 wherein said host cell is a CHO cell.

20 45. A process for producing a heterologous fusion protein comprising the steps of transcribing and translating a polynucleotide of Claim 40 under conditions wherein the heterologous fusion protein is expressed in detectable amounts.

25

46. A method for increasing neutrophil levels in a mammal comprising the administration of a therapeutically effective amount of the heterologous fusion protein of any one of Claims 27 through 36.

30

47. The use of the heterologous fusion protein as claimed in any one of Claims 22 through 29 for the manufacture of a medicament for the treatment of patients with insufficient circulating neutrophil levels.

35

48.   Use of a heterologous fusion protein of any one of
Claims 22 through 29 as a medicament.

49.   Use of a heterologous fusion protein of any one of
5   Claims 22 through 29 for the treatment of patients with
insufficient circulating neutrophil levels.

50.   A pharmaceutical formulation adapted for the treatment
of patients with insufficient neutrophil levels comprising a
10   heterologous fusion protein of any one of Claims 22 through
29.

51.   A heterologous fusion protein as hereinbefore described
with reference to any one of the Examples.
15

Abstract

The present invention encompasses heterologous fusion
proteins comprising a hyperglycsoylated G-CSF analog fused
to proteins such as albumin and the Fc portion of an
immunoglobulin which act to extend the *in vivo* half-life of
the protein compared to native G-CSF. These fusion proteins
are particularly suited for the treatment of conditions
treatable by stimulation of circulating neutrophils, such as
after chemotherapy regimens or in chronic congenital
neutropenia.

Fig. 1

1/1

```
1              5            10             15
Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys

              20           25            30
Cys Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln

         35      R1         40            45
Glu Lys Leu Cys Xaa Xaa Xaa Lys Leu Cys His Pro Glu Glu Leu Val

    50                55         R3  60      R6          65
Leu Leu Gly His Ser Leu Gly Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser
                                      R4          R7
                                  R2          R5          R8

R9          70            75            80
Xaa Xaa Xaa Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser

              85            90          R12
Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Xaa Xaa Xaa Ser Xaa
                                            R10
                                               R11

         100           105           110
Xaa Xaa Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp

         115           120           125
Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro

         130      R13   135                140  R14
Ala Leu Gln Pro Xaa Xaa Xaa Ala Met Pro Ala Phe Xaa Xaa Xaa Phe

145           150           155           160
Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Aln Ser Phe

              165           170
Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro
```

<110>   Beals, John

    Kuchibhotla, Uma

<120>   HETEROLOGOUS G-CSF FUSION PROTEINS

<130>   P-15648

<160>   66

<170>   PatentIn version 3.1

<210>   1

<211>   174

<212>   PRT

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<220>

<221>   MISC_FEATURE

<222>   (17)..(17)

<223>   Xaa at position 17 is Cys, Ala, Leu, Ser, or Glu;

<220>

<221> MISC_FEATURE

<222> (37)..(37)

<223> Xaa at position 37 is Ala or Asn;

<220>

<221> MISC_FEATURE

<222> (38)..(38)

<223> Xaa at position 38 is Thr, or any other amino acid exept Pro;

<220>

<221> MISC_FEATURE

<222> (39)..(39)

<223> Xaa at position 39 is Tyr, Thr, or Ser;

<220>

<221> MISC_FEATURE

<222> (57)..(57)

<223> Xaa at position 57 is Pro or Val;

<220>

<221> MISC_FEATURE

<222> (58)..(58)

<223> Xaa at position 58 is Trp or Asn;

<220>

<221> MISC_FEATURE

<222> (59)..(59)

<223> Xaa at position 59 is Ala or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

<222> (60)..(60)

<223> Xaa at position 60 is Pro, Thr, Asn, or Ser;

<220>

<221> MISC_FEATURE

<222> (61)..(61)

<223> Xaa at position 61 is Leu, or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

<222> (62)..(62)

<223> Xaa at position 62 is Ser or Thr;

<220>

<221> MISC_FEATURE

<222> (63)..(63)

<223> Xaa at position 63 Ser or Asn;

<220>

<221> MISC_FEATURE

<222> (64)..(64)

<223> Xaa at position 64 is Cys or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

<222> (65)..(65)

<223> Xaa at position 65 is Pro, Ser, or Thr;

<220>

<221> MISC_FEATURE

<222> (66)..(66)

<223> Xaa at position 66 is Ser or Thr;

<220>

<221> MISC_FEATURE

<222> (67)..(67)

<223> Xaa at position 67 is Gln or Asn;

<220>

<221> MISC_FEATURE

<222> (68)..(68)

<223> Xaa at position 68 is Ala or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

<222> (69)..(69)

<223> Xaa at position 69 is Leu, Thr, or Ser;

<220>

<221> MISC_FEATURE

<222> (93)..(93)

<223> Xaa at position 93 is Glu or Asn;

<220>

<221> MISC_FEATURE

<222> (94)..(94)

<223> Xaa at position 94 is Gly or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

<222> (95)..(95)

<223> Xaa at position 95 is Ile, Asn, Ser, or Thr;

<220>

<221> MISC_FEATURE

<222> (97)..(97)

<223> Xaa at position 97 is Pro, Ser, Thr, or Asn;

<220>

<221> MISC_FEATURE

<222> (133)..(133)

<223> Xaa at position 133 is Thr or Asn;

<220>

<221> MISC_FEATURE

<222> (134)..(134)

<223> Xaa at position 134 is Gln or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

<222> (135)..(135)

<223> Xaa at position 135 is Gly, Ser, or Thr;

<220>

<221> MISC_FEATURE

<222> (141)..(141)

<223> Xaa at position 141 is Ala or Asn;

<220>

<221> MISC_FEATURE

<222> (142)..(142)

<223> Xaa at position 142 is Ser or any other amino acid except Pro;

<220>

<221> MISC_FEATURE

Page 6

<222> (143)..(143)

<223> Xaa at position 143 is Ala, Ser, or Thr.

<400> 1

Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
1   5   10   15

Xaa Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
                20   25   30

Glu Lys Leu Cys Xaa Xaa Xaa Lys Leu Cys His Pro Glu Glu Leu Val
        35   40   45

Leu Leu Gly His Ser Leu Gly Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
    50   55   60

Xaa Xaa Xaa Xaa Xaa Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
65   70   75   80

Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Xaa Xaa Xaa Ser
            85   90   95

Xaa Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
            100   105   110

Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
        115   120   125

Ala Leu Gln Pro Xaa Xaa Xaa Ala Met Pro Ala Phe Xaa Xaa Xaa Phe
    130   135   140

Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
145   150   155   160

Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro
            165   170

<210> 2

<211> 1044

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 2
accccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg   60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc   120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg   180

cgtcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccaccccga ggagctggtg   240

ctcttcgaca cacggtggat gttcgacacg gtggggctcc tcgaccacct gctcggacac   300

tctctgggca tcccctgggc tcccctgagc agctgcgacg agcctgtgac agacccgtag   360

gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc   420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg   480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag   540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac   600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac   660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct   720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

aaccagaccg ccatgccggc cttcgcctct gctttccggg acgtcgggtt ggtctggcgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210> 3

<211> 1044

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 3
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cacggtggat gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

ctgggca tcccctgggc tcccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcaactct accttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agttgagatg gaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   4

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

Page 10

23> synthetic construct

<400> 4
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tcccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

Page 11

ccggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   5

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   5
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
  60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cacggtggat gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc taacactagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgat tggactcctc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc daccgtccga cgaactcggt tgaggtatcg
480

ggcctttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac

Page 12

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   6

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   6
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg ggggggacccg
 60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

tcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cacggtggat gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tcccctgagc aattgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtt aacgaccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gctggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210> 7

<211> 1044

<212> DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   7
accccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
  60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
  120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
  180

cgtcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccacccccga ggagctggtg
  240

ctcttcgaca cacggtggat gttcgacacg gtggggctcc tcgaccacct gctcggacac
  300

tctctgggca tcgttaacgc taccctgagc agctgcgacg agcctgtgac agacccgtag
  360

caattgcgat gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
  420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
  480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
  540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
  600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
  660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
  720

aaacggtggt ggtagaccgt cgtctacctt cttgacccctt accggggagc cctgcagccc
  780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
  840

Page 15

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210> 8

<211> 1044

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 8
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
  60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cacggtggat gttcgacacg gtgggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc aacgccaccc agctggcagg ctgcttgagc
420

caactccata gcgggtcgtt gcggtgggtc gaccgtccga cgaactcggt tgaggtatcg

80

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   9

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   9
accccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
60

Page 17

P-15648.ST25.txt

gacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtgccacc tacaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cacggtggat gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggcctttcc tctaccaggg gctcctgcag gccctgaacg ggacctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggacttgccc tggaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

Page 18

<210> 10

<211> 1044

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 10
accccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccacccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtgggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tcccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc daccgtccga cgaactcggt tgaggtatcg
480

ggcctttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

Page 19

acggtggt ggtagaccgt cgtctacctt cttgacccctt accggggagc cctgcagccc
780

aaccagaccg ccatgccggc cttcgcctct gctttccggg acgtcgggtt ggtctggcgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   11

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   11
accccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
 60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tcccctgagc agctgcgacg agcctgtgac agacccgtag

Page 20

60

```
gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcaactct accttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agttgagatg gaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044
```

<210> 12

<211> 1044

<212> DNA

<213> Artificial Sequence

<220>

823> synthetic construct

<400> 12

accccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcgttaacgc taccctgagc agctgcgacg agcctgtgac agacccgtag
360

caattgcgat gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag ccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

Page 22

ggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
020

cagaattccg tggaacgggt cggg
1044

<210>   13

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   13
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
  60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
  120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tcccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc aacgccaccc agctggcagg ctgcttgagc
420

caactccata gcgggtcgtt gcggtgggtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

Page 23

actgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   14

<211>   1044

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   14
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
  60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg

Page 24

240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tccctgagc agctgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtc gacgcccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gcgggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggaacg gtaccggtcc caccttggac
600

acactgcagc tggacgtcgc cgacttgcca tggccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   15

<211>   1044

<212>   DNA

Page 25

<213> Artificial Sequence

<220>

<223>   synthetic construct

<400>   15
acccccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg gggggacccg
  60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcgttaacgc taccctgagc agctgcgacg agcctgtgac agacccgtag
360

caattgcgat gggactcgtc gacgcccagc aacgccaccc agctggcagg ctgcttgagc
420

caactccata gcgggtcgtt gcggtgggtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctggaag ggatctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggaccttccc tagaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

Page 26

cggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
~~0

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210> 16

<211> 1044

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 16
accccctgg gccctgccag ctccctgccc cagagcttcc tgctcaagtg ggggacccg
   60

ggacggtcga gggacggggt ctcgaaggac gagttcgcct tagagcaagt gaggaagatc
120

cagggcgatg gcgcagcgct ccagcggaat ctcgttcact ccttctaggt cccgctaccg
180

cgtcgcgagg tcgagaagct gtgtaacacc accaagctgt gccaccccga ggagctggtg
240

ctcttcgaca cattgtggtg gttcgacacg gtggggctcc tcgaccacct gctcggacac
300

tctctgggca tcccctgggc tcccctgagc aattgcgacg agcctgtgac agacccgtag
360

gggacccgag gggactcgtt aacgaccagc caggccctgc agctggcagg ctgcttgagc
420

caactccata gctggtcggt ccgggacgtc gaccgtccga cgaactcggt tgaggtatcg
480

ggccttttcc tctaccaggg gctcctgcag gccctgaacg ggacctcccc ggaaaaggag
540

atggtccccg aggacgtccg ggacttgccc tggaggcccg agttgggtcc caccttggac
600

acactgcagc tggacgtcgc cgacgggctc aacccagggt ggaacctgtg tgacgtcgac
660

ctgcagcggc tgtttgccac caccatctgg cagcagatgg aagaactggg aatggcccct
720

aaacggtggt ggtagaccgt cgtctacctt cttgaccctt accggggagc cctgcagccc
780

acccagggtg ccatgccggc cttcgcctct gctttccggg acgtcgggtg ggtcccacgg
840

tacggccgga agcggagacg aaagcagcgc cgggcaggag gggtcctggt tgcctcccat
900

ctgcagagct tcgtcgcggc ccgtcctccc caggaccaac ggagggtaga cgtctcgaag
960

ctggaggtgt cgtaccgcgt cttaaggcac cttgcccagc ccgacctcca cagcatggcg
1020

cagaattccg tggaacgggt cggg
1044

<210>   17

<211>   1762

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   17
gatgcgcaca agagtgaggt tgctcatcgg tttaaagatt tgggagaaga aaatttcaaa
  60

gccttggtgt tgattgcctt tgctcagtat cttcagcagt gtccatttga agatcatgta

aaattagtga atgaagtaac tgaatttgca aaaacatgtg ttgctgatga gtcagctgaa
180

aattgtgaca aatcacttca tacccttttt ggagacaaat tatgcacagt tgcaactctt
240

cgtgaaacct atggtgaaat ggctgactgc tgtgcaaaac aagaacctga gagaaatgaa
300

tgcttcttgc aacacaaaga tgacaaccca aacctccccc gattggtgag accagaggtt
360

gatgtgatgt gcactgcttt tcatgacaat gaagagacat ttttgaaaaa atacttatat
420

gaaattgcca gaagacatcc ttacttttat gccccggaac tccttttctt tgctaaaagg
480

tataaagctg cttttacaga atgttgccaa gctgctgata aagctgcctg cctgttgcca
540

aagctcgatg aacttcggga tgaagggaag gcttcgtctg ccaaacagag actcaagtgt
600

gccagtctcc aaaaatttgg agaaagagct ttcaaagcat gggcagtagc tcgcctgagc
660

cagagatttc ccaaagctga gtttgcagaa gtttccaagt tagtgacaga tcttaccaaa
720

gtccacacgg aatgctgcca tggagatctg cttgaatgtg ctgatgacag ggcggacctt
780

gccaagtata tctgtgaaaa tcaagattcg atctccagta aactgaagga atgctgtgaa
840

aaacctctgt tggaaaaatc ccactgcatt gccgaagtgg aaaatgatga gatgcctgct
900

gacttgcctt cattagctgc tgattttgtt gaaagtaagg atgtttgcaa aaactatgct
960

gaggcaaagg atgtcttcct gggcatgttt ttgtatgaat atgcaagaag gcatcctgat
1020

tactctgtcg tgctgctgct gagacttgcc aagacatatg aaaccactct agagaagtgc
1080

tgtgccgctg cagatcctca tgaatgctat gccaaagtgt tcgatgaatt taaacctctt

40

gtggaagagc ctcagaattt aatcaaacaa aattgtgagc tttttgagca gcttggagag
1200

tacaaattcc agaatgcgct attagttcgt tacaccaaga aagtacccca agtgtcaact
1260

ccaactcttg tagaggtctc aagaaaccta ggaaaagtgg gcagcaaatg ttgtaaacat
1320

cctgaagcaa aaagaatgcc ctgtgcagaa gactatctat ccgtggtcct gaaccagtta
1380

tgtgtgttgc atgagaaaac gccagtaagt gacagagtca ccaaatgctg cacagaatcc
1440

ttggtgaaca ggcgaccatg cttttcagct ctggaagtcg atgaaacata cgttcccaaa
1500

gagtttaatg ctgaaacatt caccttccat gcagatatat gcacactttc tgagaaggag
1560

agacaaatca agaaacaaac tgcacttgtt gagctcgtga aacacaagcc caaggcaaca
1620

aaagagcaac tgaaagctgt tatggatgat ttcgcagctt ttgtagagaa gtgctgcaag
1680

gctgacgata aggagacctg ctttgccgag gagggtaaaa aacttgttgc tgcaagtcaa
1740

gctgccttag gcttataatg ac
1762

<210> 18

<211> 232

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 18

Page 30

Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro
1   5   10   15

Ala Pro Glu Lys Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro
        20   25   30

Lys Asp Thr Lys Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val
        35   40   45

Val Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val
    50   55   60

Asp Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln
65   70   75   80

Tyr Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln
                85   90   95

Asp Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala
            100   105   110

Leu Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro
        115   120   125

Arg Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr
    130   135   140

Lys Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser
145   150   155   160

Asp Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr
            165   170   175

Lys Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr
        180   185   190

Ser Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe
    195   200   205

Page 31

```
Ser Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys
    210             215             220

Ser Leu Ser Leu Ser Pro Gly Lys
225             230
```

<210> 19

<211> 229

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic contruct

<400> 19

```
Glu Ser Lys Tyr Gly Pro Pro Cys Pro Pro Cys Pro Ala Pro Glu Phe
1               5               10              15

Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr
            20              25              30

Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val
            35              40              45

Ser Gln Glu Asp Pro Glu Val Gln Phe Asn Trp Tyr Val Asp Gly Val
    50              55              60

Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Phe Asn Ser
65              70              75              80

Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu
                85              90                  95

Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Gly Leu Pro Ser
```

```
                    100                  105                  110
·Ser Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro
         115                  120                  125

Gln Val Tyr Thr Leu Pro Pro Ser Gln Glu Glu Met Thr Lys Asn Gln
     130                  135                  140

Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala
145                  150                  155                  160

Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr
                 165                  170                  175

Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Arg Leu
             180                  185                  190

Thr Val Asp Lys Ser Arg Trp Gln Glu Gly Asn Val Phe Ser Cys Ser
         195                  200                  205

Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser
     210                  215                  220

Leu Ser Leu Gly Lys
225
```

<210> 20

<211> 585

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 20

Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu
                  5   10   15

Glu Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln
              20   25   30

Gln Cys Pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu
          35   40   45

Phe Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys
      50   55   60

Ser Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu
65   70   75   80

Arg Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro
              85   90   95

Glu Arg Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu
              100   105   110

Pro Arg Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His
          115   120   125

Asp Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala Arg
      130   135   140

Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg
145   150   155   160

Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala
              165   170   175

Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser
          180   185   190

Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu
          195   200   205

Page 34

Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro
    210   215   220

Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys
225   230   235   240

Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp
            245   250   255

Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser
        260   265   270

Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His
    275   280   285

Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser
    290   295   300

Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala
305   310   315   320

Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg
            325   330   335

Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr
        340   345   350

Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu
    355   360   365

Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro
370   375   380

Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Asn Leu Gly Glu
385   390   395   400

Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro
        405   410   415

Page 35

Gln Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys
        420   425   430

Val Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys
        435   440   445

Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His
    450   455   460

Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser
465   470   475   480

Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr
                485   490   495

Tyr Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp
            500   505   510

Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala
        515   520   525

Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu
    530   535   540

Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys
545   550   555   560

Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val
            565   570   575

Ala Ala Ser Gln Ala Ala Leu Gly Leu
            580   585

<210>   21

<211>   703

<212>   DNA

Page 36

13> Artificial Sequence

<220>

<223> synthetic construct

<400> 21
gagcccaaat cttgtgacaa aactcacaca tgcccaccgt gcccagcacc tgaactcctg
60

gggggaccgt cagtcttcct cttcccccca aaacccaagg acaccctcat gatctcccgg
120

acccctgagg tcacatgcgt ggtggtggac gtgagccacg aagaccctga ggtcaagttc
180

aactggtacg tggacggcgt ggaggtgcat aatgccaaga caaagccgcg ggaggagcag
240

tacaacagca cgtaccgtgt ggtcagcgtc ctcaccgtcc tgcaccagga ctggctgaat
300

ggcaaggagt acaagtgcaa ggtctccaac aaagccctcc cagcccccat cgagaaaacc
360

atctccaaag ccaaagggca gccccgagaa ccacaggtgt acaccctgcc cccatcccgg
420

gaggagatga ccaagaacca ggtcagcctg acctgcctgg tcaaaggctt ctatcccagc
480

gacatcgccg tggagtggga gagcaatggg cagccggaga caactacaa gaccacgcct
540

cccgtgctgg actccgacgg ctccttcttc ctctatagca agctcaccgt ggacaagagc
600

aggtggcagc aggggaacgt cttctcatgc tccgtgatgc atgaggctct gcacaaccac
660

tacacgcaga agagcctctc cctgtctccg ggtaaatgat agt
703

<210> 22

<211> 981

Page 37

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 22
tccaccaagg gcccatcggt cttccgcta gcgccctgct ccaggagcac ctccgagagc
60

acagccgccc tgggctgcct ggtcaaggac tacttccccg aaccggtgac ggtgtcgtgg
120

aactcaggcg ccctgaccag cggcgtgcac accttcccgg ctgtcctaca gtcctcagga
180

ctctactccc tcagcagcgt ggtgaccgtg ccctccagca gcttgggcac gaagacctac
240

acctgcaacg tagatcacaa gcccagcaac accaaggtgg acaagagagt tgagtccaaa
300

tatggtcccc catgcccacc ctgcccagca cctgagttcc tggggggacc atcagtcttc
360

ctgttccccc caaaacccaa ggacactctc atgatctccc ggacccctga ggtcacgtgc
420

gtggtggtgg acgtgagcca ggaagacccc gaggtccagt tcaactggta cgtggatggc
480

gtggaggtgc ataatgccaa gacaaagccg cgggaggagc agttcaacag cacgtaccgt
540

gtggtcagcg tcctcaccgt cctgcaccag gactggctga acggcaagga gtacaagtgc
600

aaggtctcca acaaaggcct cccgtcctcc atcgagaaaa ccatctccaa agccaaaggg
660

cagccccgag agccacaggt gtacaccctg cccccatccc aggaggagat gaccaagaac
720

caggtcagcc tgacctgcct ggtcaaaggc ttctacccca gcgacatcgc cgtggagtgg
780

gagagcaatg ggcagccgga gaacaactac aagaccacgc ctcccgtgct ggactccgac

Page 38

ggctccttct tcctctacag caggctaacc gtggacaaga gcaggtggca ggaggggaat
900

gtcttctcat gctccgtgat gcatgaggct ctgcacaacc actacacaca gaagagcctc
960

tccctgtctc tgggtaaatg a
981

<210> 23

<211> 406

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 23

Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
1   5   10   15

Ala Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
            20   25   30

Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu Val
        35   40   45

Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser Cys
    50   55   60

Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
65   70   75   80

Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile Ser
                85   90   95

Page 39

Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
                  100   105   110

Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
              115   120   125

Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe Ala Ser Ala Phe
        130   135   140

Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
145   150   155   160

Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro Glu Pro
                165   170   175

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu
              180   185   190

Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp
        195   200   205

Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp
    210   215   220

Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly
225   230   235   240

Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn
              245   250   255

Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp
        260   265   270

Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro
        275   280   285

Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu
290   295   300

Page 40

Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr Lys Asn
305   310   315   320

Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile
                325   330   335

Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr
            340   345   350

Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys
        355   360   365

Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys
    370   375   380

Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu
385   390   395   400

Ser Leu Ser Pro Gly Lys
            405

<210>   24

<211>   403

<212>   PRT

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   24

Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
1   5   10   15

Ala Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln

Page 41

```
          20                    25                    30

Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu Val
    35                  40                  45

Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser Cys
    50                  55                  60

Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
65                  70                  75                  80

Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile Ser
            85                  90                  95

Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
            100                 105                 110

Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
        115                 120                 125

Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe Ala Ser Ala Phe
    130                 135                 140

Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
145                 150                 155                 160

Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro Glu Ser
            165                 170                 175

Lys Tyr Gly Pro Pro Cys Pro Ser Cys Pro Ala Pro Glu Phe Leu Gly
            180                 185                 190

Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu Met
            195                 200                 205

Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser Gln
    210                 215                 220
```

Page 42

Glu Asp Pro Glu Val Gln Phe Asn Trp Tyr Val Asp Gly Val Glu Val
225   230   235   240

His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Phe Asn Ser Thr Tyr
                245   250   255

Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn Gly
                260   265   270

Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Gly Leu Pro Ser Ser Ile
                275   280   285

Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro Gln Val
                290   295   300

Tyr Thr Leu Pro Pro Ser Gln Glu Glu Met Thr Lys Asn Gln Val Ser
305   310   315   320

Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val Glu
                325   330   335

Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr Pro Pro
                340   345   350

Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Arg Leu Thr Val
                355   360   365

Asp Lys Ser Arg Trp Gln Glu Gly Asn Val Phe Ser Cys Ser Val Met
370   375   380

His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser Leu Ser
385   390   395   400

Leu Gly Lys

<210>   25

<211>   500

12> PRT

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 25

Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
1   5   10   15

Ala Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
                20   25   30

Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu Val
        35   40   45

Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser Cys
    50   55   60

Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
65   70   75   80

Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile Ser
                85   90   95

Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
            100   105   110

Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
        115   120   125

Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe Ala Ser Ala Phe
    130   135   140

Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
145   150   155   160

Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro Gly Gly
                165   170   175

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ala His
            180   185   190

Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe
        195   200   205

Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro
    210   215   220

Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu Phe Ala Lys
225   230   235   240

Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys Ser Leu His
            245   250   255

Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu Arg Glu Thr
            260   265   270

Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn
        275   280   285

Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu
    290   295   300

Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His Asp Asn Glu
305   310   315   320

Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro
            325   330   335

Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala
            340   345   350

Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu

Page 45

```
               355                   360                   365


    Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys
        370                 375                 380


    Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe
    385                 390                 395                 400


    Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu
                    405                 410                 415


    Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys Val His Thr
                    420                 425                 430


    Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp
                435                 440                 445


    Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu
        450                 455                 460


    Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His Cys Ile Ala
    465                 470                 475                 480


    Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser Leu Ala Ala
                    485                 490                 495


    Asp Phe Val Glu
                500


    <210>   26

    <211>   69

    <212>   DNA

    <213>   Artificial Sequence


    <220>
```

Page 46

<223> synthetic construct

<400> 26
gtaagcttgc gtcgacgcta gcggcgcgcc gccatggccg gacctgccac ccagagcccc
60

atgaagctg
69

<210> 27

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 27
ggggcaggga gctggctggg cccagtggag tggcttcctg cactgtccag agtgcactgt
60

g
61

<210> 28

<211> 59

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 28
ggacagtgca ggaagccact ccactgggcc cagccagctc cctgccccag agcttcctg
59

<210> 29

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 29
gaacctcgag gatcctcatt agggctgggc aaggtgcctt aagacgcggt acgacacctc
60

caggaagctc tg
72

<210> 30

<211> 69

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 30
gtaagcttgc gtcgacgcta gcggcgcgcc gccatggccg gacctgccac ccagagcccc
60

atgaagctg
69

<210> 31

<211> 57

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 31
gctctaaggc cttgagcagg aagctctggg gcagggagct cgctgggccc agtggag
 57

<210> 32

<211> 53

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 32
gggcccagcg agctccctgc cccagagctt cctgctcaag gccttagagc aag
 53

<210> 33

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 33
gaacctcgag gatcctcatt agggctgggc aaggtgcctt aagacgcggt acgacacctc
 60

caggaagctc tg
 72

<210> 34

<211> 69

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 34
gtaagcttgc gtcgacgcta gcggcgcgcc gccatggccg gacctgccac ccagagcccc
 60

atgaagctg
 69

<210> 35

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 35
gtccgagcag cactagttcc tcggggtggc acagcttggt ggtgttacac agcttctcct
 60

g
 61

<210> 36

<211> 66

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 36
ggcgcagcgc tccaggagaa gctgtgtaac accaccaagc tgtgccaccc cgaggaacta
 60

gtgctg
 66

<210> 37

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 37
gaacctcgag gatcctcatt agggctgggc aaggtgcctt aagacgcggt acgacacctc
 60

caggaagctc tg
 72

<210> 38

<211> 69

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 38
gtaagcttgc gtcgacgcta gcggcgcgcc gccatggccg gacctgccac ccagagcccc
 60

atgaagctg
 69

<210> 39

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 39
gcccggcgct ggaaagcgct ggcgaaggcc ggcatggcgg tctggttggg ctgcagggca
 60

g
 61

<210> 40

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 40
ggcccctgcc ctgcagccca accagaccgc catgccggcc ttcgccagcg ctttccagcg
 60

<210> 41

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 41
gaacctcgag gatcctcatt agggctgggc aaggtgcctt aagacgcggt acgacacctc   60

caggaagctc tg   72

<210> 42

<211> 69

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 42
gtaagcttgc gtcgacgcta gcggcgcgcc gccatggccg gacctgccac ccagagcccc   60

atgaagctg   69

<210> 43

<211> 68

<212> DNA

<213> Artificial Sequence

<220>

<223>   synthetic construct

<400>   43
gcccggcgct ggaaggtaga gttgaaggcc ggcatggcac cctgggtggg ctgaagagca
 60

ggggccat
 68

<210>   44

<211>   74

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

<400>   44
gggaatggcc cctgctcttc agcccaccca gggtgccatg ccggccttca actctacctt
 60

ccagcgccgg gcag
 74

<210>   45

<211>   72

<212>   DNA

<213>   Artificial Sequence

<220>

<223>   synthetic construct

gaacctcgag gatcctcatt agggctgggc aaggtgcctt aagacgcggt acgacacctc
   60 .

caggaagctc tg
   72

<210> 46

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 46
gctagcggcg cgccaccatg
   20

<210> 47

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 47
gctcagggta gcgttaacga tgcccagaga gtg
   33

<210> 48

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 48
gggcatcgtt aacgctaccc tgagcagctg
 30

<210> 49

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 49
gactcgagga tcctcattag ggctggg
 27

<210> 50

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 50
gctagcggcg cgccaccatg gccggacctg ccacccag
 38

<210> 51

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 51
caagcagccg gccagctggg tggcgttgct ggggcagctg ctcag   45

<210> 52

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 52
gccccagcaa cgccacccag ctggccggct gcttgag   37

<210> 53

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

23> synthetic construct

<400> 53
gactcgagga tcctcattag ggctgggcaa ggtgccttaa gacgcgg
 47

<210> 54

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 54
gctagcggcg cgccaccatg
 20

<210> 55

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 55
ggggcaacta gtcaggttag cccaggg
 27

<210> 56

<211> 27

<212> DNA

13> Artificial Sequence

<220>

<223> synthetic construct

<400> 56
gctaacctga ctagttgccc cagccag
 27

<210> 57

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 57
gactcgagga tcctcattag ggctggg
 27

<210> 58

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 58
gctagcggcg cgccaccatg
 20

Page 59

<210> 59

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 59
ggtgcaattg ctcaggggag cccag
 25

<210> 60

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 60
gcaattgcac cagccaggcc ctg
 23

<210> 61

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<210> 61

gactcgagga tcctcattag ggctggg
27

<210> 62

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 62
gctagcggcg cgccaccatg gccggacctg ccacccag
38

<210> 63

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 63
ccggactggt cccgttcagg gcctgcagga gcccctg
37

<210> 64

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 64
gaacgggacc agtccggagt tgggtcccac cttgg
 35

<210> 65

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 65
gactcgagga tcctcattag ggctgggcaa ggtgccttaa gacgcgg
 47

<210> 66

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic construct

<400> 66
gtcgacgcta gcggcgcgcc accatggccg gacctg
 36